

АНАЛІЗ СУЧАСНИХ ТЕНДЕНЦІЙ ТА ПІДХОДІВ ДО НАДІЙНОГО І ЗАХИЩЕНОГО ЗБЕРІГАННЯ ВЕЛИКИХ ДАНИХ

О. Р. Дейнека

Національний університет “Львівська політехніка”,
кафедра захисту інформації
E-mail: oleh.r.deineka@lpnu.ua

© Дейнека О. Р., 2025

Стрімке накопичення інформаційних активів потребує нових підходів до їх зберігання та захисту. Стаття присвячена аналізу сучасних підходів до зберігання великих обсягів даних із врахуванням їхньої ефективності, надійності та безпеки. Розглянуто ключові технології, такі як хмарні платформи, локальні рішення та розподілені системи зберігання, а також особливості їхнього застосування.

У дослідженні підкреслено роль хмарних технологій як основи для зберігання й обробки великих даних, що забезпечують масштабованість, гнучкість і високу доступність. Обговорено виклики, пов'язані з конфіденційністю, залежністю від постачальників послуг і адаптацією технологій до специфічних потреб організацій.

Проведено детальне порівняння локальних і хмарних рішень для зберігання даних з точки зору вартості, доступності, продуктивності та безпеки. Наведені результати спрямовані на розробку більш ефективних і надійних рішень для управління великими даними в умовах зростаючих кіберзагроз і динамічного розвитку технологій.

Ключові слова: великі дані, зберігання, захист даних, конфіденційність, цілісність, доступність, хмарні технології, локальні технології зберігання, машинне навчання.

Вступ

Сучасний світ характеризується стрімким накопиченням інформаційних активів, значна частина яких містить критично важливу інформацію. Збільшення обсягів таких даних висуває вимоги до їх класифікації за різноманітними параметрами, пошуків методів та способів надійного зберігання, ефективної передачі, а також захисту від несанкціонованого доступу. Зростання кількості атак на інформаційні ресурси потребує постійного вдосконалення стандартів, підходів і методів у сфері кібербезпеки. Це передбачає як розробку нових механізмів протидії зловмисним діям, так і вдосконалення інфраструктури для захисту даних.

Окрему увагу приділяють розробці стандартів безпечного зберігання інформації, які дають змогу забезпечити контроль доступу, підтримувати конфіденційність і гарантувати захищеність даних у різних умовах. Стандарти безпеки варіюються залежно від країни, галузі, рівня конфіденційності інформації та специфічних вимог організацій. Часто ці стандарти запроваджуються з використанням міжнародних рекомендацій, за участю незалежних аудиторських компаній, які підтверджують відповідність встановленим нормам.

Попри наявність численних інструментів для організації роботи з великими обсягами даних, професіонали стикаються з низкою викликів. До основних проблем належать забезпечення цілісності даних, конфіденційності та доступності. Одним із найбільш складних завдань є збереження незмінності інформації на всіх етапах її життєвого циклу – від моменту створення до пошуку та аналізу. Не менш важливою є проблема конфіденційності, яка потребує, щоб доступ до інформації мали лише авторизовані особи. Крім того, стрімке зростання обсягів даних ускладнює забезпечення швидкого доступу до потрібної інформації, що є критичним для ефективного функціонування сучасних організацій.

Хмарні технології стали однією з ключових інновацій у сфері зберігання великих обсягів даних. Вони забезпечують високу масштабованість, гнучкість і доступність, даючи змогу організаціям ефективно управляти даними незалежно від їхнього обсягу. Завдяки моделям “інфраструктура як послуга” (IaaS) та “платформа як послуга” (PaaS) компанії отримують можливість зберігати й обробляти дані без значних витрат на фізичне обладнання. Хмарні платформи також пропонують інтегровані інструменти для забезпечення безпеки даних, включно з шифруванням, багаторівневим доступом і резервним копіюванням. Водночас використання хмарних рішень спричиняє низку викликів, таких як залежність від постачальників послуг, питання конфіденційності та дотримання регуляторних норм. Проте хмарні технології залишаються одним із найбільш перспективних напрямів для надійного зберігання даних у сучасному цифровому середовищі.

Огляд літературних джерел

Великі дані – це об’ємні та різномірні набори інформації, які створюються у великих кількостях і надходять із різних джерел. Вони можуть містити дані про клієнтів, товари й послуги, результати опитувань, аналіз вебтрафіку та багато іншого, що накопичується практично в кожній організації. Такі дані мають значний потенціал для підтримки прийняття рішень, оптимізації процесів і підвищення ефективності роботи організації.

Для вирішення завдань, пов’язаних зі зберіганням і обробкою великих даних, існують численні комерційні рішення. Проте такі рішення часто є дорогими й потребують складної інтеграції, налаштування та постійного технічного супроводу. Це спонукає наукову спільноту та галузь до активних досліджень і розробок нових методів і підходів у сфері зберігання великих даних.

Попри велику кількість досліджень [1-4] та запропонованих підходів, не всі вони відповідають ключовим вимогам до надійності, безпеки та масштабованості. Ті рішення, які забезпечують високий рівень захисту даних, можуть бути недостатньо універсальними для різних типів інформації або не відповідати специфічним потребам окремих організацій. У більшості випадків організації орієнтуються на міжнародні стандарти безпеки, перевірені незалежними аудиторськими компаніями. Однак навіть за дотримання цих стандартів залишаються виклики, пов’язані з адаптацією підходів до динамічного характеру сучасних великих даних, що потребує подальших досліджень і вдосконалення практичних рішень.

Великі дані [5-6] та хмарні обчислення [7-8] є ключовими технологіями, які займають центральне місце в сучасній інформаційній екосистемі. Щодня з різноманітних джерел, таких як соціальні мережі, датчики IoT, фінансові транзакції та інші джерела, генерується колосальний обсяг даних. Ці дані характеризуються не лише значними обсягами, а й високою швидкістю їхнього надходження та різноманітністю форматів, що ускладнює їхнє опрацювання за допомогою традиційних інструментів. Концепція великих даних охоплює всі аспекти роботи з такими наборами даних: від їхнього зберігання до аналізу, який дає змогу отримувати корисну інформацію [9]. Хмарні обчислення забезпечують необхідну інфраструктуру для обробки цих даних, пропонуючи рентабельні, масштабовані та ефективні рішення. Поєднання великих даних із хмарними технологіями дає змогу організаціям ефективно вирішувати завдання, пов’язані з аналізом даних, знижуючи витрати на локальну інфраструктуру.

Сьогодні багато галузей, зокрема бізнес, охорона здоров'я, освіта, логістика та державне управління, активно інтегрують можливості великих даних у свої процеси. Наприклад, у медицині аналітика великих даних допомагає в прогнозуванні епідемій, персоналізованому лікуванні та покращенні ефективності систем охорони здоров'я. Водночас хмарні обчислення надають інструменти для обробки цих даних у режимі реального часу, що є критично важливим для прийняття швидких і обґрунтованих рішень. Це поєднання технологій відкриває нові перспективи для розвитку інновацій у різних секторах економіки.

Постановка задачі

Зберігання великих обсягів даних є однією з ключових задач сучасних інформаційних систем. З розвитком технологій та зростанням кількості даних, які генеруються щодня, виникає необхідність у розробці нових підходів та методів для ефективного та безпечного зберігання інформації.

Метою дослідження є аналіз сучасних підходів до зберігання великих обсягів даних із урахуванням їх ефективності, надійності та безпеки. Дослідження спрямоване на вивчення основних технологій, таких як хмарні платформи, локальні рішення та розподілені системи зберігання, із оцінкою їхніх переваг і недоліків.

Особливу увагу варто приділити питанням забезпечення захищеності даних, які є ключовим аспектом у сучасному цифровому середовищі. У межах роботи також необхідно розглянути виклики, з якими стикаються організації під час впровадження нових технологій зберігання, зокрема питання конфіденційності, цілісності та доступності даних. Результати дослідження мали б сприяти розвитку більш надійних підходів до управління даними та забезпечення їх безпеки в умовах стрімкого зростання обсягів інформації.

Хмарні сервіси для зберігання великих даних

Хмарне сховище є ключовою частиною зберігання та роботи з великими даними. Використання хмарних технологій для роботи з великими даними передбачає зберігання та аналіз об'ємних наборів даних за допомогою хмарних сервісів. Завдяки високій ефективності та гнучкості таких рішень дедалі більше компаній інтегрують хмарні платформи для управління своєю інформацією. Щороку інвестиції бізнесу в хмарні сервіси продовжують зростати, що підтверджує їхню важливість у сучасному цифровому середовищі. Хмарні технології дають змогу організаціям зберігати дані на віддалених серверах, що забезпечує доступ до інформації з будь-якої точки світу.

Перехід до хмарних рішень для великих даних зумовлений *кількома ключовими факторами*:

- по-перше, стрімке збільшення обсягів даних та висока вартість локальної інфраструктури змушують організації шукати більш економічно вигідні й масштабовані альтернативи;
- по-друге, компанії стикаються з потребою обробляти інформацію з різноманітних джерел, таких як CRM-системи, пристрої IoT, фінансові транзакції тощо;
- по-третє, гнучкість хмарних рішень дає змогу швидко реагувати на непередбачувані коливання даних, забезпечуючи масштабованість і надійність у реальному часі.

Крім того, зростає попит на оперативну аналітику, яка допомагає компаніям залишатися конкурентоспроможними в умовах динамічного ринку. Розвиток віддаленої роботи також став важливим стимулом для впровадження хмарних платформ, адже вони забезпечують доступ до даних з будь-якої точки світу. Отже, хмарні технології не лише оптимізують управління великими даними, а й створюють передумови для підвищення ефективності бізнесу та адаптації до сучасних викликів.

Основні переваги хмарних рішень:

- масштабованість, яка дає змогу організаціям легко збільшувати або зменшувати обсяги зберігання відповідно до потреб, що особливо важливо в умовах швидкого зростання даних;

– гнучкість, яка дає змогу адаптуватися до змін у бізнес-процесах та вимогах, забезпечуючи швидкий доступ до необхідних ресурсів;

– економія витрат на інфраструктуру досягається завдяки відсутності необхідності у придбанні та обслуговуванні власного обладнання.

Проте існують також певні недоліки хмарних рішень. Залежність від інтернет-з'єднання може стати критичною у випадку перебоїв у роботі мережі, що може призвести до тимчасової недоступності даних. Крім того, питання конфіденційності даних у хмарних середовищах залишається актуальним, оскільки дані можуть бути доступні третім сторонам. Важливо також враховувати можливі ризики безпеки, пов'язані з використанням хмарних сервісів, такі як несанкціонований доступ до даних або їх втрата.

Хмарні платформи для роботи з великими даними вирізняються своєю ефективністю завдяки **п'яти ключовим аспектам:**

1. *Зберігання.* Хмарні платформи забезпечують масштабоване сховище даних завдяки використанню розподілених файлових систем. Це дає змогу зберігати великі обсяги структурованих і неструктурованих даних без ризику браку місця, що особливо важливо в умовах стрімкого зростання інформації.

2. *Обробка.* Однією з головних переваг хмарних платформ є швидкість обробки даних. Завдяки використанню таких інструментів, як Hadoop і Spark, платформи здатні одночасно обробляти величезні обсяги інформації на численних вузлах, що значно скорочує час аналізу.

3. *Інтеграція.* Хмарні рішення забезпечують простоту інтеграції за допомогою API, які дають змогу легко поєднувати дані з різних джерел як внутрішніх, так і зовнішніх. Крім того, хмарні провайдери пропонують набір попередньо інтегрованих сервісів, які допомагають прискорити впровадження систем.

4. *Аналітика.* Хмарні платформи часто містять інструменти машинного навчання (ML) і штучного інтелекту (AI), що дають змогу автоматизувати аналіз даних і виявлення шаблонів. Такі сервіси, як Amazon SageMaker або Google Cloud Vertex AI, дають змогу створювати, навчати й розгортати ML-моделі для більш глибокого та точного аналізу.

5. *Безпека.* Хоча деякі організації можуть сумніватися в безпечності хмари, сучасні хмарні платформи пропонують високий рівень захисту даних. Це включає потужні механізми шифрування, багаторівневий контроль доступу, регулярні перевірки систем безпеки та швидке оновлення захисних інструментів у відповідь на нові загрози.

Крім того, хмарні платформи є гнучким рішенням, яке дає змогу організаціям адаптуватися до зростаючих потреб у зберіганні й обробці даних. Вони сприяють оптимізації бізнес-процесів, знижуючи витрати на інфраструктуру та технічне обслуговування, водночас забезпечуючи високий рівень надійності й доступності. Усе це робить хмарні технології невід'ємною складовою сучасного управління великими даними.

Однак успіх платформ великих даних у хмарі багато в чому залежить від правильного архітектурного підходу. Розглянемо різні хмарні сервіси, які дають можливість зберігати велику кількість даних.

1. Amazon Web Services (AWS):

– Amazon S3 (Simple Storage Service)

Amazon S3 є одним з найпопулярніших сервісів для зберігання об'єктів, що забезпечує високу доступність та надійність даних. S3 дає змогу зберігати будь-яку кількість даних у вигляді об'єктів, які організовані в бакети. Кожен об'єкт складається з даних, метаданих та унікального ідентифікатора. S3 підтримує різні класи зберігання, такі як S3 Standard, S3 Intelligent-Tiering, S3 Standard-IA (Infrequent Access), S3 One Zone-IA та S3 Glacier, що дає змогу оптимізувати витрати на зберігання залежно від частоти доступу до даних.

Безпека даних в Amazon S3 забезпечується за допомогою різних механізмів, таких як шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (IAM), а також підтримка політик доступу на рівні бакетів та об'єктів. S3 також підтримує журналювання доступу за допомогою AWS CloudTrail, що дає змогу відстежувати всі операції з даними [10].

– *Amazon EBS (Elastic Block Store)*

Amazon EBS надає можливість створення блочних сховищ для використання з віртуальними машинами EC2. EBS забезпечує високу продуктивність та масштабованість, даючи змогу створювати томи різних типів, таких як General Purpose SSD (gp2), Provisioned IOPS SSD (io1), Throughput Optimized HDD (st1) та Cold HDD (sc1). EBS також підтримує функції резервного копіювання та відновлення даних за допомогою знімків (snapshots).

Безпека даних в Amazon EBS забезпечується за допомогою шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (IAM), а також підтримка політик доступу на рівні томів. EBS також підтримує журналювання доступу за допомогою AWS CloudTrail, що дає змогу відстежувати всі операції з даними [11].

– *Amazon Glacier*

Amazon Glacier призначений для довгострокового зберігання архівних даних з низькою частотою доступу. Glacier забезпечує низькі витрати на зберігання та високу надійність даних, даючи змогу зберігати великі обсяги інформації протягом тривалого часу. Glacier підтримує різні режими відновлення даних, такі як Expedited, Standard та Bulk, що дозволяє вибирати оптимальний баланс між вартістю та швидкістю відновлення.

Безпека даних в Amazon Glacier забезпечується за допомогою шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (IAM), а також підтримка політик доступу на рівні архівів. Glacier також підтримує журналювання доступу за допомогою AWS CloudTrail, що дає змогу відстежувати всі операції з даними [12].

2. Microsoft Azure:

– *Azure Blob Storage*

Azure Blob Storage забезпечує зберігання неструктурованих даних, таких як документи, зображення та відео, з високою доступністю та надійністю. Blob Storage підтримує три типи об'єктів: блокові блоби (block blobs), сторінкові блоби (page blobs) та додаткові блоби (append blobs). Azure Blob Storage також підтримує різні рівні доступу, такі як Hot, Cool та Archive, що дає змогу оптимізувати витрати на зберігання залежно від частоти доступу до даних.

Безпека даних в Azure Blob Storage забезпечується за допомогою шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (RBAC), а також підтримка політик доступу на рівні контейнерів та об'єктів. Blob Storage також підтримує журналювання доступу за допомогою Azure Monitor, що дозволяє відстежувати всі операції з даними [13].

– *Azure Disk Storage*

Azure Disk Storage надає можливість створення блочних сховищ для використання з віртуальними машинами. Azure Disk Storage підтримує різні типи дисків, такі як Standard HDD, Standard SSD, Premium SSD та Ultra Disk, що дає змогу вибирати оптимальний баланс між продуктивністю та вартістю. Azure Disk Storage також підтримує функції резервного копіювання та відновлення даних за допомогою знімків (snapshots).

Безпека даних в Azure Disk Storage забезпечується за допомогою шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (RBAC), а також підтримка політик доступу на рівні дисків. Disk Storage також підтримує журналювання доступу за допомогою Azure Monitor, що дає змогу відстежувати всі операції з даними [14].

– *Azure Archive Storage*

Azure Archive Storage призначений для довгострокового зберігання архівних даних з низькою частотою доступу. Archive Storage забезпечує низькі витрати на зберігання та високу надійність

даних, даючи змогу зберігати великі обсяги інформації протягом тривалого часу. Archive Storage підтримує функції резервного копіювання та відновлення даних, що дозволяє забезпечити високу надійність та доступність інформації.

Безпека даних в Azure Archive Storage забезпечується за допомогою шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (RBAC), а також підтримка політик доступу на рівні контейнерів та об'єктів. Archive Storage також підтримує журналювання доступу за допомогою Azure Monitor, що дає змогу відстежувати всі операції з даними [15].

3. Google Cloud Platform (GCP)

– Google Cloud Storage

Google Cloud Storage забезпечує зберігання об'єктів з високою доступністю та надійністю, що дає змогу зберігати великі обсяги даних з можливістю швидкого доступу. Cloud Storage підтримує різні класи зберігання, такі як Standard, Nearline, Coldline та Archive, що дозволяє оптимізувати витрати на зберігання залежно від частоти доступу до даних. Google Cloud Storage також підтримує функції резервного копіювання та відновлення даних, що дозволяє забезпечити високу надійність та доступність інформації.

Безпека даних в Google Cloud Storage забезпечується за допомогою шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (IAM), а також підтримка політик доступу на рівні бакетів та об'єктів. Cloud Storage також підтримує журналювання доступу за допомогою Google Cloud Audit Logs, що дозволяє відстежувати всі операції з даними [16].

– Google Persistent Disk

Google Persistent Disk надає можливість створення блочних сховищ для використання з віртуальними машинами. Persistent Disk підтримує різні типи дисків, такі як Standard Persistent Disks, SSD Persistent Disks та Balanced Persistent Disks, що дає змогу вибирати оптимальний баланс між продуктивністю та вартістю. Persistent Disk також підтримує функції резервного копіювання та відновлення даних за допомогою знімків (snapshots).

Безпека даних в Google Persistent Disk забезпечується за допомогою шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (IAM), а також підтримка політик доступу на рівні дисків. Persistent Disk також підтримує журналювання доступу за допомогою Google Cloud Audit Logs, що дозволяє відстежувати всі операції з даними [17].

– Google Archive Storage

Google Archive Storage призначений для довгострокового зберігання архівних даних з низькою частотою доступу. Archive Storage забезпечує низькі витрати на зберігання та високу надійність даних, даючи змогу зберігати великі обсяги інформації протягом тривалого часу. Archive Storage підтримує функції резервного копіювання та відновлення даних, що дозволяє забезпечити високу надійність та доступність інформації.

Безпека даних в Google Archive Storage забезпечується за допомогою шифрування даних під час зберігання та передачі, контроль доступу на основі ролей (IAM), а також підтримка політик доступу на рівні бакетів та об'єктів. Archive Storage також підтримує журналювання доступу за допомогою Google Cloud Audit Logs, що дає змогу відстежувати всі операції з даними [18].

Загалом дані сервіси для зберігання даних від великих постачальників хмарних платформ мають низку ключових характеристик, що робить їх привабливими для організацій, які мають справу з великими обсягами даних. Вони надають гнучкість у виборі типів зберігання, рівнів доступу, а також механізми забезпечення безпеки, які відповідають різним вимогам щодо продуктивності та витрат. У табл. 1 ми порівнюємо ці сервіси за кількома основними параметрами. Це дасть змогу організаціям вибирати оптимальне рішення залежно від їхніх потреб у зберіганні, обробці та захисті даних, враховуючи специфіку бізнес-процесів та бюджету.

Таблиця 1

Порівняння хмарних сервісів для зберігання великих даних

Параметр	Amazon S3	Azure Blob Storage	Google Cloud Storage
Класи зберігання	Standard, Intelligent-Tiering, Standard-AI, One Zone-IA, Glacier	Hot, Cool, Archive	Standard, Nearline, Coldline, Archive
Доступність	99.99 % (Standard)	99.99 % (Hot)	99.95 % (Standard)
Вартість зберігання (за ГБ/місяць)	\$0.023(Standard)	\$0.0184 (Hot)	\$ 0.020 (Standard)
Вартість запитів	\$0.005 за 1000 запитів PUT,COPY,POST,LIST	\$0.004 за 10000 Put запитів	\$0.005 за 1000 Class A запитів
Вартість передачі даних	\$0.09 за за ГБ (перші 10ТБ/місяць)	\$0.087 за ГБ (перші 50 ТБ/місяць)	\$0.12 за ГБ (перші 1 ТБ/місяць)
Безпека даних	Шифрування, IAM, CloudTrail	Шифрування, RBAC, Azure Monitor	Шифрування IAM

Локальні технології зберігання даних

Локальне зберігання даних [19] передбачає використання фізичних носіїв інформації, таких як жорсткі диски або флеш-пам'ять, розміщених на серверному обладнанні, яке належить організації. Всі аспекти організації зберігання даних, включаючи структуру файлів, визначаються внутрішніми вимогами організації, що надає їй повний контроль над інформацією. Організація також несе відповідальність за забезпечення безпеки та захист від несанкціонованого доступу до даних.

Локальні технології зберігання, на відміну від хмарних рішень, забезпечують автономність і можливість налаштувати інфраструктуру відповідно до специфічних потреб підприємства. Це дає змогу здійснювати більш детальне управління доступом до інформації, зокрема створення дерев файлів, обмеження за квотами даних та надання індивідуальних дозволів користувачам. Однак, незважаючи на ці переваги, локальні рішення мають і свої недоліки, зокрема високі витрати на обладнання та постійну необхідність у технічному обслуговуванні.

Локальні сервери зберігання даних також обмежені в мобільності, оскільки доступ до інформації можливий лише в межах фізичної локації організації, що може стати проблемою в умовах зростаючої потреби в віддаленому доступі. Проте для підприємств, які мають специфічні вимоги до продуктивності, безпеки та контролю доступу, локальні технології зберігання є найбільш доцільним вибором. Типові приклади локальних рішень: файлові сервери, мережеві накопичувачі (NAS), жорсткі диски та флеш-накопичувачі, які активно використовують як великі компанії, так і окремі користувачі для зберігання та обробки даних.

Розглянемо системи, які можна розгорнути на локальних серверах:

– Однією з найпоширеніших систем для локального зберігання даних є Hadoop Distributed File System (HDFS). HDFS є розподіленою файловою системою, яка забезпечує високу доступність та надійність даних. Вона дає змогу зберігати великі обсяги даних на кластері серверів, розподіляючи їх на блоки та дублюючи для забезпечення стійкості до збоїв. HDFS широко використовується в аналітичних системах та обробці великих даних [20]. HDFS також підтримує масштабованість, що дозволяє системі ефективно обробляти і зберігати дані в умовах постійного зростання обсягів інформації.

– Іншою популярною системою є Network Attached Storage (NAS). NAS - спеціалізований пристрій для зберігання даних, який під'єднується до мережі та забезпечує доступ до файлів для користувачів та додатків. NAS системи можуть бути масштабованими та підтримувати різні рівні RAID для забезпечення надійності та продуктивності.

– Також варто згадати про сервери з Direct Attached Storage (DAS). DAS є простим та ефективним рішенням для зберігання даних, яке передбачає під'єднання дисків безпосередньо до сервера. DAS системи можуть бути використані для невеликих організацій або для специфічних задач, які не потребують складної інфраструктури. DAS також дає змогу досягати високої швидкості доступу до даних завдяки прямому під'єднанню до сервера, що робить його оптимальним для обробки великих обсягів інформації в реальному часі [22- 23]. Однак, незважаючи на свою простоту та ефективність, DAS обмежена в масштабованості та не забезпечує того рівня гнучкості, який пропонують інші системи зберігання, такі як NAS або SAN.

Забезпечення безпеки даних на локальних серверах

Забезпечення безпеки даних на локальних серверах є критично важливим аспектом для будь-якої організації. Основні методи забезпечення безпеки передбачають:

– Шифрування даних дає змогу захистити інформацію від несанкціонованого доступу. Шифрування може бути реалізоване як на рівні дисків, так і на рівні файлів. Використання сучасних алгоритмів шифрування, таких як AES-256, забезпечує високий рівень захисту даних, а також дозволяє зберігати дані у захищеному вигляді навіть у разі фізичного доступу до носіїв інформації.

– Контроль доступу дає змогу визначати права користувачів на доступ до даних. Використання ролей та політик доступу допомагає забезпечити, що тільки авторизовані користувачі можуть отримати доступ до конфіденційної інформації. Інтеграція з системами аутентифікації, такими як Active Directory, дає змогу централізовано керувати доступом до даних та підтримувати відповідність нормативним вимогам щодо контролю доступу.

– Резервне копіювання є важливим елементом забезпечення безпеки даних. Регулярне створення резервних копій дає змогу відновити дані у випадку збоїв або атак. Використання різних методів резервного копіювання, таких як повне, інкрементальне та диференційне, допомагає забезпечити високу надійність та ефективність процесу, а також гарантує швидке відновлення в разі втрати інформації.

– Моніторинг та журналювання дають змогу відстежувати всі операції з даними та виявляти потенційні загрози. Використання систем моніторингу, таких як SIEM (Security Information and Event Management), дозволяє централізовано збирати та аналізувати журнали подій, що допомагає виявляти та реагувати на інциденти безпеки, а також забезпечує прозорість операцій з даними для виявлення потенційних порушень безпеки.

Переваги локальних рішень

Основні переваги локальних рішень - повний контроль над даними, висока безпека, незалежність від зовнішніх постачальників послуг та можливість налаштування інфраструктури відповідно до специфічних потреб організації.

Повний контроль над даними дає змогу організаціям самостійно визначати політики доступу, шифрування та резервного копіювання. Це забезпечує високий рівень конфіденційності та захисту інформації.

Висока безпека досягається завдяки використанню сучасних методів захисту, таких як шифрування, контроль доступу та моніторинг. Локальні рішення дають змогу організаціям самостійно контролювати всі аспекти безпеки даних, що знижує ризики несанкціонованого доступу або втрати інформації.

Незалежність від зовнішніх постачальників послуг дає змогу організаціям уникнути залежності від інтернет-з'єднання та забезпечити безперервний доступ до даних навіть у випадку перебоїв у роботі мережі. Це особливо важливо для критично важливих додатків, які потребують високої доступності та надійності.

Можливість налаштування інфраструктури відповідно до специфічних потреб організації дає змогу оптимізувати витрати та забезпечити високу продуктивність. Локальні рішення дозволяють організаціям вибирати обладнання та програмне забезпечення, яке найкраще відповідає їхнім вимогам та бюджету.

Однак локальні технології також мають свої недоліки. Висока вартість обладнання та необхідність постійного технічного обслуговування можуть стати значним фінансовим навантаженням для організацій. Крім того, забезпечення резервного копіювання та відновлення даних може потребувати значних ресурсів та часу.

Локальні технології зберігання даних, на відміну від хмарних, передбачають зберігання інформації на власних серверах організації. Це дозволяє мати повний контроль над даними та забезпечувати високий рівень безпеки. Основні переваги локальних рішень - незалежність від зовнішніх постачальників послуг та можливість налаштування інфраструктури відповідно до специфічних потреб організації. Однак локальні технології також мають свої недоліки - це висока вартість обладнання та необхідність постійного технічного обслуговування.

Крім того, локальні рішення дають змогу організаціям забезпечити високий рівень кастомізації в розгортанні інфраструктури, що може бути критично важливо для підприємств з унікальними вимогами. Вони також дозволяють зберігати більшу частину операцій всередині організації, що може зменшити залежність від сторонніх постачальників і підвищити рівень контролю за власними даними та процесами.

Таблиця 2

Порівняння локальних технологій для зберігання великих даних

Параметр	HDFS (Hadoop Distributed File System)	Nas(Network Attached Storage)	Das (Direct Attached Storage)
Тип зберігання	Розподілена файлова система	Файлова система	Блочне зберігання
Основне призначення	Обробка та зберігання великих даних	Централізоване зберігання файлів	Локальне зберігання даних
Масштабованість	Висока	Середня	Низька
Продуктивність	Висока	Середня	Висока
Надійність	Висока	Висока (RAID)	Залежить від конфігурації
Вартість обслуговування	Висока	Середня	Низька
Контроль доступу	ACL, Kerberos	ACL,LDAP,AD	ACL
Шифрування даних	Підтримується	Підтримується	Підтримується
Резервне копіювання	Підтримується	Підтримується	Підтримується
Моніторинг	Підтримується (Hadoop Metrics)	Підтримується	Залежить від конфігурації
Основні компоненти	NameNode, DataNode	NAS сервери, RAID масиви	Диски, контролери
Переваги	Висока масштабованість, продуктивність	Простота використання, централізація	Висока продуктивність, низька затримка
Недоліки	Висока вартість, складність налаштування	Обмежена Масштабованість	Обмежена масштабованість, залежить від сервера

Результати дослідження

Хмарне та локальне сховище мають істотні відмінності, що стосуються власності, доступу до даних та їх зберігання. Власник і контроль над інфраструктурою є ключовими аспектами, які розрізняють ці два типи сховищ. Хмарні провайдери управляють зберіганням даних та доступом до

них, а також забезпечують можливість масштабування системи відповідно до потреб бізнесу. Натомість локальне сховище зазвичай розташоване на конкретному пристрої або сервері, який безпосередньо керує даними на місці.

Розглянемо питання безпеки даних у хмарних та локальних системах. Хмарне сховище пропонує зручність підписки без необхідності витрат на обслуговування апаратного забезпечення. Замість значних одноразових витрат на фізичні носії даних, хмарні провайдери пропонують чітку цінову модель через щомісячні або річні платежі. Однак для користувачів, які віддають перевагу контролю над своїми даними, локальне сховище може бути вигіднішим, оскільки ви не залежите від сторонніх серверів і можете зберігати дані на фізичних носіях.

З погляду доступності, хмарне сховище виграє завдяки тому, що дозволяє доступ до даних з будь-якої точки світу за умови наявності інтернету. Однак в разі проблем із під'єднанням до інтернету або низькою швидкістю з'єднання це може стати обмеженням. Локальне сховище не має подібних проблем, оскільки можна працювати з даними без необхідності підключатися до інтернету.

Що стосується конфіденційності, хмарне сховище зазвичай надає високий рівень безпеки завдяки шифруванню та іншим заходам, однак дані, що зберігаються на віддалених серверах, можуть бути більш вразливими до атак з боку хакерів. Локальне сховище дає змогу тримати дані під повним контролем, що дає можливість забезпечити вищий рівень конфіденційності.

Окрім того, локальне сховище потребує регулярних витрат на обслуговування і оновлення обладнання, що може стати значним фінансовим навантаженням, тоді як хмарні сервіси надають гнучкість і зручність, даючи змогу бізнесу зосередитись на основних завданнях, а не на технічних питаннях інфраструктури.

Хмарні платформи ідеально підходять для організацій, які прагнуть знизити початкові витрати та швидко розширювати свої ресурси відповідно до змінних потреб бізнесу. Однак важливо ретельно налаштувати політики доступу та безпеки.

Таблиця 3

Порівняння локальних та хмарних сервісів для зберігання великих даних

Параметр	Локальні системи (HDFS, NAS,DAS)	Хмарні системи (AWS,Azure,GCP)
1	2	3
Ціна	Висока початкова вартість обладнання та обслуговування. Постійні витрати на технічне обслуговування та оновлення	Гнучка модель оплати. Низька початкова вартість, але постійні витрати на використання ресурсів
Безпека	Високий рівень контролю над безпекою. Можливість впровадження власних політик безпеки. Захист даних на фізичному рівні	Високий рівень безпеки, забезпечений постачальниками послуг. Підтримка шифрування, контроль доступу та відповідність стандартам безпеки (GDPR, HIPAA)
Масштабованість	Обмежена масштабованість. Потребує додаткових витрат на розширення інфраструктури	Висока масштабованість. Можливість швидкого збільшення або зменшення ресурсів відповідно до потреб
Продуктивність	Висока продуктивність для локальних задач. Низька підтримка доступу до даних	Висока продуктивність, але залежить від інтернет-з'єднання. Можливість затримки під час передачі великих обсягів даних
Надійність	Висока надійність за правильної конфігурації. Потребує резервного копіювання та відновлення даних	Висока надійність, забезпечена постачальниками послуг. Автоматичне резервне копіювання та відновлення даних

Продовження табл. 3

1	2	3
Контроль доступу	Повний контроль доступу до даних. Можливість налаштування політик доступу відповідно до потреб організації	Контроль доступу забезпечується постачальником послуг. Підтримка ролей та політик (IAM, RBAC)
Шифрування даних	Підтримується на рівні дисків та файлів. Можливість використання власних алгоритмів шифрування	Підтримується на рівні зберігання та передачі даних. Використання сучасних алгоритмів шифрування (AES-256)
Резервне копіювання	Потребує власних рішень для резервного копіювання. Можливість налаштування частоти та методів резервного копіювання	Автоматичне резервне копіювання, забезпечене постачальником послуг. Можливість налаштування політик резервного копіювання
Моніторинг	Потребує власних рішень для моніторингу та журналювання. Можливість використання SIEM систем	Вбудовані інструменти моніторингу та журналювання (Cloud Watch, Azure Monitor, Stackdriver)
Гнучкість	Висока гнучкість у налаштуванні та управлінні інфраструктурою. Можливість адаптації до специфічних потреб організації	Висока гнучкість у використанні ресурсів. Можливість швидкого масштабування та адаптації до змін у бізнес-процесах
Залежність від інтернету	Незалежність від інтернет-з'єднання. Безперервний доступ до даних навіть у випадку перебоїв у роботі мережі	Залежить від інтернет-з'єднання. Можливі перебої у доступі до даних у разі проблем з мережею
Відповідність стандартам	Можливість самостійного забезпечення відповідності стандартам та регуляціям	Відповідність стандартам та регуляціям забезпечується постачальником послуг (GDPR, HIPAA, ISO 27001)
Технічна підтримка	Потребує власної команди технічної підтримки. Можливість швидкого реагування на потреби	Технічна підтримка забезпечується постачальником послуг. Різні рівні підтримки залежно від обраного плану

Висновки

У сучасному цифровому середовищі ефективно зберігання та захист великих обсягів даних є критично важливими завданнями для організацій. Проведено всебічний аналіз сучасних підходів до зберігання великих обсягів даних, зокрема хмарних і локальних технологій.

Хмарні рішення виявилися перспективними завдяки високій масштабованості, економічності та гнучкості в управлінні даними, особливо у поєднанні з механізмами захисту, такими як шифрування, багаторівневий контроль доступу та резервне копіювання. Водночас вони супроводжуються викликами, зокрема залежністю від постачальників, ризиками конфіденційності та складністю дотримання регуляторних вимог.

Локальні технології, такі як HDFS, NAS і DAS, забезпечують високий рівень автономності, контроль над даними та надійність, проте потребують значних початкових витрат і постійного технічного обслуговування.

Результати дослідження підкреслюють, що вибір між хмарними й локальними рішеннями має базуватися на специфічних потребах, ресурсах і пріоритетах організації. Особливу увагу потрібно приділяти інтеграції новітніх технологій, таких як блокчейн і розподілені системи, для підвищення надійності, прозорості та безпеки. Це допоможе організаціям адаптуватися до швидко змінюваного інформаційного ландшафту та забезпечити надійний захист своїх інформаційних активів.

Список літератури

1. Aujla G. S., Chaudhary R., Kumar N., Das A. K., Rodrigues J. J. *SecSVA: secure storage, verification, and auditing of big data in the cloud environment*. *IEEE Commun Mag.* 2018. 56(1). 78-85. DOI: 10.1109/MCOM.2018.1700379.
2. Deineka O., Harasymchuk O., Partyka A., Obshta A., Korshun N. *Designing Data Classification and Secure Store Policy According to SOC 2 Type II*. *CEUR Workshop Proceedings*. 2024. 3654. Pp. 398–409.
3. Yatskiv V., Kulyna S., Yatskiv N., Kulyna H. *Protected Distributed Data Storage Based on Residue Number System and Cloud Services*. *10th International Conference on Advanced Computer Information Technologies (ACIT)*, 2020. 796–799. DOI: 10.1109/acit49673.2020.9208849.
4. Kai Z. *Research on network data storage Technology based on Autonomous Controllable system*. In *2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)*, Shanghai, China, 2021. 183- 186. DOI: 10.1109/icceai52939.2021.00035.
5. Jambi Kamal, Eassa Fathy, Alshomrani Abdullah. *A Framework to Secure Big Data Storage*. *Journal of Computational and Theoretical Nanoscience*. 2017. 14. 5600- 5605. DOI: 10.1166/jctn.2017.6990.
6. Deibe David, Amor Margarita, Doallo Ramon. *Big data storage technologies: a case study for web-based LiDAR visualization*. *IEEE International Conference on Big Data*. 2018. Pp. 3831–3840. DOI: 10.1109/BigData.2018.8622589.
7. Reena M., Nargunam A. S. *Secured Storage of Big Data in Cloud*. *International Journal of Recent Technology and Engineering (IJRTE)*. 2019. 8(2S3). Pp. 6- 10. DOI: 10.35940/ijrte.B1002.0782S319.
8. Martseniuk Y., Partyka A., Harasymchuk O., Korshun N. *Automated Conformity Verification Concept for Cloud Security Cybersecurity Providing in Information and Telecommunication Systems 2024*, 3654. Pp. 25- 37.
9. Коваль. О., Гарасимчук О. *Розроблення гібридного методу побудови сховища даних*. *CSN*. 2024. *Bun.* 6. № 1. С. 67–80. DOI:10.23939/csn2024.01.067.
10. Amazon Web Services, Inc. *Amazon S3 Documentation*. 2023. URL: <https://docs.aws.amazon.com/s3/index.html>.
11. Amazon Web Services, Inc. *Amazon EBS Documentation*. 2023. URL: <https://docs.aws.amazon.com/ebs/index.html>.
12. Amazon Web Services, Inc. *Amazon Glacier Documentation*. 2023. URL: <https://docs.aws.amazon.com/glacier/index.html>.
13. Microsoft Corporation. *Azure Blob Storage Documentation*. 2023. URL: <https://docs.microsoft.com/en-us/azure/storage/blobs/>.
14. Microsoft Corporation. *Azure Disk Storage Documentation*. 2023. URL: <https://docs.microsoft.com/en-us/azure/virtual-machines/disks-types>.
15. Microsoft Corporation. *Azure Archive Storage Documentation*. 2023. URL: <https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blob-archive>.
16. Google LLC. *Google Cloud Storage Documentation*. 2023. URL: <https://cloud.google.com/storage/docs/>.
17. Google LLC. *Google Persistent Disk Documentation*. 2023. URL: <https://cloud.google.com/persistent-disk/docs/>.
18. Google LLC. *Google Archive Storage Documentation*. 2023. URL: <https://cloud.google.com/storage/docs/storage-classes#archive>.
19. Гарасимчук О. Бужович О. *Стратегії та інноваційні підходи до захисту баз даних в епоху зростаючих кіберзагроз. Безпека інформації*. 2024. Т. 30. № 1. С. 166- 178. DOI: 10.18372/2225-5036.30.18618.
20. Дейнека О., Гарасимчук О. *Дослідження проблем класифікації та безпечно зберігання даних. Безпека інформації*. 2023. 29. № 3. С. 147- 153. DOI: 10.18372/2225-5036.29.18074.
21. White Tom. *Hadoop: The Definitive Guide*. O'Reilly Media. 2015. 756 p. ISBN: 978-1491901632.
22. Silberschatz Abraham, Galvin Peter B., Gagne Greg. *Operating System Concepts*. Wiley. 2018. 976 p. ISBN: 978-1119456339.
23. Barker Richard, Paul Massiglia. *Storage Area Network Essentials: A Complete Guide to Understanding and Implementing SANs*. Wiley. 2001. 480 p. ISBN: 978-0471034450.

**ANALYSIS OF CURRENT TRENDS AND APPROACHES TO RELIABLE
AND SECURE BIG DATA STORAGE****O. R. Deineka**

Lviv Polytechnic National University,
Department of Information Protection

© *Deineka O. R.*, 2025

The rapid accumulation of information assets requires new approaches to their storage and protection. The article is devoted to the analysis of modern approaches to storing large volumes of data, taking into account their efficiency, reliability and security. Key technologies such as cloud platforms, local solutions and distributed storage systems are considered, as well as the features of their application.

The study emphasizes the role of cloud technologies as a basis for storing and processing big data, ensuring scalability, flexibility and high availability. The challenges associated with confidentiality, dependence on service providers and adaptation of technologies to the specific needs of organizations are discussed.

A detailed comparison of local and cloud solutions for data storage in terms of cost, availability, performance and security is presented. The results of the development are presented for the development of more effective and reliable solutions for managing big data in the context of growing cyber threats and dynamic development of technologies.

Keywords: big data, storage, data protection, confidentiality, integrity, availability, cloud technologies, local storage technologies, machine learning.