МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Наукові праці
Донецького національного технічного університету

Серія: “Інформатика, кібернетика та обчислювальна техніка”

Всеукраїнський науковий збірник

Заснований у травні 1996 року

Виходить 2 рази на рік

№ 2(33)’ 2021 - №1(34)’ 2022

Покровськ – 2022

УДК 004.3+004.9+004.2+51.7+519.6+519.7

Публікується згідно з рішенням Вченої ради ДВНЗ «Донецький національний технічний університет» (протокол № 4 від 24.05.2022).

Збірник містить наукові статті співробітників ДонНТУ та інших навчальних і наукових закладів України, які є науковими партнерами ДонНТУ. Публікації висвітлюють результати наукових досліджень і розробок в таких напрямках, як інформатика, чисельні методи, паралельні обчислення, програмування, розробка засобів обчислювальної техніки, дослідження комп'ютерних мереж, машинна графіка і обробка зображень, математичне моделювання в різних галузях. Матеріали збірника призначені для наукових співробітників, викладачів, інженерно-технічних працівників, аспірантів та студентів.

Засновник та видавець – Донецький національний технічний університет (ДонНТУ)

Адреса редакції:
Юридична адреса: 85300, Україна, Донецька область, м. Покровськ, пл. Шибанкова, 2, ДВНЗ «ДонНТУ», 3.312.
Фактична адреса: 43012, Україна, Волинська область, м. Луцьк, вул. Софії Ковалевської, 29, ДВНЗ «ДонНТУ», 219.
E-mail: yevhen.bashkov@donntu.edu.ua

Збірник зареєстровано в Державному комітеті інформаційної політики, телебачення та радіомовлення України. Свідоцтво: серія КВ, №7374 від 03.06.2003.

Збірник включено до переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії за спеціальностями 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 123 Комп'ютерна інженерія (наказ Міністерства освіти і науки України №409 від 17 березня 2020 р.)

Збірник "Наукові праці ДонНТУ, серія "Інформатика, кібернетика та обчислювальна техніка" за наказом № 409 МОНУ від 17. 03.2020 отримав категорію Б.

ISSN 1996-1588

ЗМІСТ

УДК 004.056

[1]Н. Маслова, канд. техн. наук, доц.,
[2]А. Нікітенко, магістр
Донецький національний технічний університет, м. Покровськ, Україна
[1]nataliia.maslova@donntu.edu.ua
[2]andrii.nikitenko.kita@donntu.edu.ua

Особливості захисту даних великих обсягів на online ресурсах

Великі дані – один з викликів сучасності, пов'язаний з необхідністю обробки зростаючих обсягів даних. Тенденцією при роботі з великими даними є застосування online – ресурсів. Найбільш важливим аспектом при роботі у хмарному середовищі є забезпечення на віддаленому ресурсі безпеки даних - комп'ютерної, мережевої, інформаційної. У роботі пропонується методика й програмний інструмент розміщення даних у хмарних середовищах, базовою умовою створення яких є вимоги дотримання критеріїв безпеки

Ключові слова: Великі дані, інформаційна безпека, хмарні середовища, online ресурси.

DOI: 10.31474/1996-1588-2021-2-33-24-32

Вступ

Big Data, або великі дані – це об'ємні набори структурованих, неструктурованих та напівструктурованих категорій даних. Ці набори досить великі й різноманітні, їх складно або навіть неможливо обробити традиційними засобами роботи з даними. Незважаючи на те, що сучасна реальність прийому, передачі, обробки та створення сховищ пов'язана з вимогами ретельного ставлення до безпеки, конфіденційності та цілісності даних, кожний провайдер, орієнтований на роботу з Big Data пропонує своє рішення. Це різноманіття приводить до необхідності проведення аналізу ринку послуг, обрання або побудови методології розміщення великих даних, й ця задача є актуальною в умовах постійного зростання обсягів даних та кіберзлочинності.

Мета дослідження: аналіз механізмів забезпечення захисту великих даних на online ресурсах та розробка методики розміщення даних великих обсягів на online-ресурсах різних моделей розгортання й опис її реалізації.

Завданнями дослідження є:

- аналіз механізмів забезпечення захисту великих даних на online-ресурсах;
- визначення критеріїв обрання об'єктних хмарних сховищ з точки зору гарантування відповідності критеріям безпеки;
- опис методики розміщення даних великих обсягів та додатку для розбивки, розподілу та реплікації даних на online-ресурсах.

Проблеми захисту Big Data

Основними характеристиками великих даних є обсяг, різноманітність, швидкість генерації та накопичення, вразливість, достовірність та складність. У англомовній літературі це терміни Volume, Variety, Velocity, Vulnerability, Veracity та Complexity [1].

В даний час значна кількість даних генерується з різних джерел, включаючи сайти соціальних мереж, різні віддалені датчики, сигнали GPS стільникового телефону, записи транзакцій та файли журналів. Завдяки Інтернету щодня виробляється близько терабайту даних, і більша частина цієї інформації має невід'ємні бізнес-цінності. Крім того, дані можуть містити персональну інформацію, й проблема захисту доповнюється питаннями конфіденційності.

Проблеми захисту даних великих обсягів виявилися достатньо складними, й термін Vulnerability (вразливість) додано в перелік характеристик Big Data як один з найважливіших.

Забезпеченням захисту великих даних займаються такі компанії, як Cloud Security Alliance (CSA), Національний інститут стандартів і технологій США (NIST), Агентство Європейського Союзу з питань мережевої та інформаційної безпеки (European Union Agency for Network and Information Security, ENISA), й, безумовно, американська компанія International Business Machines (IBM), Microsoft Corporation, Oracle Corporation та технологічна компанія Hewlett Packard [2].

Стан забезпечення захисту великих даних та його особливості ми маємо можливість спостерігати та аналізувати за щорічно поновлюваним оглядам компанії Arcserve (Мінесота, США) [3].

На сьогоднішній день актуальними проблеми, пов'язаними з захистом великих даних є витік даних, неправильна конфігурація та неадекватний контроль змін, відсутність архітектури та стратегії хмарної безпеки, недостатня ідентичність та управління ключами, викрадення облікового запису, інсайдерська загроза, небезпечні інтерфейси користувача (UI) та API.

По даним на грудень 2020 року виділяють

такі обов'язкові до застосування заходи з організації захисту даних: шифрування; резервне копіювання та відновлення даних; дотримання всіх стандартів відповідності, оцінювання стану безпеки та надання рекомендацій щодо проведення подальших робіт з захисту даних.

В найбільш узагальненому вигляді спеціалісти рекомендують зосередитися на чотирьох напрямках захисту великих даних:

- безпеці інфраструктури;
- криптографічному захисті;
- забезпеченню конфіденційності даних;
- процедурах керування даними.

Все перелічене повинно підтримуватися процедурами безперервного моніторингу безпеки.

Технології та платформи роботи з великими даними

Безпека інфраструктури. Інфраструктура складається з апаратного забезпечення, програмного забезпечення, мереж та засобів, на яких працюють хмарні сервіси.

Найбільш розповсюдженою практикою роботи з великими даними є застосування хмарних, або online – ресурсів. Значну увагу створенню та застосуванню технологій роботи з великими даними приділяють відомі постачальники on-line послуг: Amazon, Microsoft і Google.

Через важливість на сьогоднішній день безпеки в області хмарних обчислень, перераховані вище провайдери не тільки надають користувачам послуги з накопичення, обробки та збереження даних значних обсягів, але й забезпечують безпеку прийому-передачі цих даних, застосування надійних систем та платформ їх обробки.

Серед значної кількості пропозицій щодо технологій обробки великих даних основними інструментами та технологіями роботи з великими даними є бібліотеки масово-паралельної обробки, системи управління базами даних категорії NoSQL, технології Hadoop.

Як показує аналіз інтернет-джерел [4-6], у якості базової платформи й Amazon (AWS S3), Microsof Azure й Google Cloud Platform застосовують Hadoop.

Основною перевагою платформи, яка приваблює інтернет-гігантів є те, що платформа працює ефективніше класичних реляційних баз даних. Крім того, перевагами платформи є розподілена файлова система та близько тисячі різноманітних сервісів для обробки інформації, які доповнюють Hadoop.

На попередніх етапах недоліками Hadoop вважалися складність платформи та недостатній рівень безпеки, але сучасна версія Apache Hadoop 3.3.0 підтримує ARM-архітектуру, Java 11, систему каталогів YARN-додатків, файлову систему Tencent Cloud COS для доступу до об'єктного сховища COS і планування запуску контейнерів за розкладом. Анонсовано полегшення роботи з DNS і IP, а також стабілізація HDFS RBF (Router-based Federation). У версії 3.3.0 вдосконалено засоби керування безпекою, а саме [7]:

– підтримку безпеки маршрутизатором HDFS;

- запобігання атак і несанкціонованого доступу до Big Data ззовні;

- забезпечення безпеки використання великих даних внутрішніми клієнтами (у тому числі й автоматизованими системами);

- жорстке адміністрування та моніторинг всіх завдань, пов'язаних з безпекою Big Data.

Прикладами таких рішень є використання та настроювання загальнодоступних апаратно-програмних рішень Apache Knox Gateway, Apache Ranger та Atlas. Навіть невеликі та середні підприємства мають можливість розгорнути власну, приватну хмару, у якості платформи обробки застосувати Hadoop, застосувати згадані вище інструменти посилення захисту й додати інші можливості, у тому числі, кодування, контроль доступу, резервне копіювання.

Загальний вигляд однієї з схем захисту рішення Hadoop наведено на рисунку 1 й описано, у тому числі, у роботі [8].

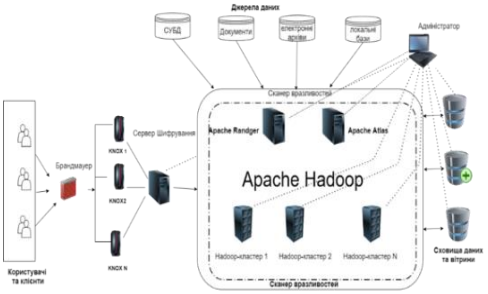

Рисунок 1 – Схема захисту Hadoop- рішення

В останній версії Apache Hadoop проведено значну роботу з підсилення безпеки, Cloud-гіганти застосовують платформу для зберігання й обробки великих даних, але доповнюють її власними інструментами та алгоритмами сучасного захисту.

Криптографічний захист великих даних

Згідно методології NIST Big Data Security and Privacy [9] для криптографічного захисту великих даних в хмарних сховищах рекомендується використовувати одну з технологій: гомоморфне шифрування, функціональне шифрування, шифрування на основі політик контроля доступу та технології Blockchain (рис. 2).

Повне гомоморфне шифрування – модель шифрування, яка дозволяє довільне різноманіття додавання та множення, і, отже, виконує безліч видів обчислень без потреби в дешифруванні

Функціональне шифрування – це тип шифрування з відкритим ключем, в якому власник секретного ключа може визначити функцію, за допомогою якої отримано шифр-текст.

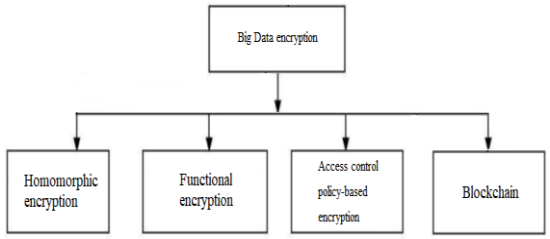

Рисунок 2 – Рекомендації NIST

Blockchain – розподілена база даних, що зберігає впорядкований ланцюжок записів (блоків). Дані користувача перетворюються в блок, кожен блок в мережі має унікальне хеш-значення, що й забезпечує захист даних.

Якщо ж перейти безпосередньо до розгляду алгоритмів, які застосовуються при захисті великих даних , то це:

- варіації алгоритму AES, який відносять до розділу криптографії із секретним ключем;
- алгоритм RSA, який є одним із перших успішних криптографічних алгоритмів криптографії з публічним ключем;
- гібридна криптографія, яка поєднує переваги криптографічних алгоритмів з відкритим ключем із продуктивністю криптографії із секретним ключем;
- алгоритми нейронної криптографії, принцип роботи яких полягає в імпульсивній провідності нейронів;
- сучасні алгоритми квантової криптографії, засновані на принципах невизначеності Гейзенберга, поляризації та заплутаності фотонів.

Методика розподілу та розміщення даних

Останнім часом багато великих компаній або підприємств зосереджуються на проблемах, пов'язаних з обробкою та обслуговуванням їх даних, ці компанії чи підприємства часто витрачають багато часу та накладних витрат на керування даними, але наслідки іноді розчаровують. Це особливо вірно, коли ці компанії чи підприємства не мають великої ділової кореляції з управлінням даними, у такій ситуації ці компанії чи підприємства хотіли б знайти надійного та професійного постачальника, щоб мати справу зі своїми даними, при збереженні, керування та обслуговування, а не робити це самостійно. Звичайно, що передача великих даних і їх віддалене зберігання можуть спричинити багато проблем, наприклад, занадто довгий час обробки, час від часу система виходить з ладу та можуть виникати деякі невизначені системні збої. Передача великих даних безпосередньо в центр обробки даних хмарного сховища може бути вразливою до збою системи через величезні розміри даних.

Альтернативним шляхом забезпечення конфіденційності і захисту великих даних, є розподіл даних на декількох хмарах.

Ідея розподілу даних на декількох хмарах не є новою.

Так, особливістю методики, пропонованої у роботі [10] є можливість поділу масиву даних значного обсягу на блоки, кількість яких залежить від кількості доступних хмар. Замість шифрування автори пропонують використовувати фрагментацію. Реплікація або інші засоби для збереження доступності відсутні.

Автори P. Suwansrikham та K. She у роботі [11] також не акцентують увагу на криптографічному захисті, й вважають, що застосування великої кількості різноманітних хмар для розміщення даних автоматично додає гарантії їх захищеності.

Більш ретельний підхід пропонують автори роботи [12]. Згідно цієї методики паралельно з діленням файлів виконується реплікація даних та розподіл по хмарах. Проводиться шифрування шляху до файлу з використанням односторонньої функції з секретом. Ділити файл можливо в залежності від кількості доступних хмар

Наприкінці обзору, слід згадати про методику, опубліковану у [13]. В ній автори запропонували власний алгоритм шифрування. Ця методика потребує проведення процедури автентифікації та отримання дозволу на доступ від власника даних.

Таким чином, ідея розподілу даних значних обсягів між декількома хмарами не є принципово новою, але не є стандартизованою та доказовою.

Тому у даній роботі пропонується авторська методика, яка враховує переваги та недоліки методик, наведених у роботах [10-13].

Методика, що за назвою програмного продукту отримала умовну назву SkyProtect потребує проведення процедури автентифікації та передбачає можливість додаткової реплікації шифрованих даних на приватне хмарне сховище. Передбачено шифрування даних з застосуванням алгоритму AES-128 на стороні користувача. Розподіл даних проводиться з урахуванням гарантій забезпечення безпеки даних, котрі надаються тим чи іншим постачальником хмарних послуг з урахуванням ступеню важливості даних.

Так, у запропонованій схемі дані великих обсягів будуть розділені на менші блоки даних, і ці менші блоки даних будуть зберігатися в хмарних носіях даних один за іншим. Оскільки ці блоки даних набагато менші, ніж загальні масиви, вони достатньо ефективно можуть бути захищені відомими методами шифрування в процесі віддаленої передачі та зберігання.

При завантаженні даних великих обсягів потрібно враховувати рівень важливості вхідних даних. Поділимо їх на 3 рівня.

Перший рівень - без додаткового захисту – дані, що не містять у собі конфіденційних (або важливих) даних, захист покладаємо на хмарних провайдерів, перелік яких складається з урахуванням задекларованих володарями сервісів можливостей.

Другий рівень - з додатковим захистом (шифруванням). Це дані, що містять конфіденційну інформацію. Перед завантаженням на хмари вона шифрується та розподіляється серед хмарних провайдерів.

Третій рівень з додатковим захистом (шифруванням) та реплікацією даних. Це дані, що містять не тільки конфіденційну, але й критичну інформацію, перед завантаженням на хмари вона шифрується, розподіляється серед хмарних провайдерів та додатково реплікується на приватну хмару.

Декілька рівнів важливості даних великих обсягів було додано через необхідність мінімізації витрат на завантаження не критичної та не конфіденційної інформації на об'єктні хмарні сховища та підвищення ефективності самого програмного забезпечення.

Критерії вибору online ресурсів

С писок критично важливих критеріїв вибору об'єктних хмарних сховищ був розроблений з точки зору відповідності критеріям захисту. У якості таких критеріїв обрано:

- максимальний розмір об'єкта, що завантажується – наскільки великі дані можна завантажити в контейнер за раз;
- ліміт на кількість об'єктів в контейнері – в якій кількості можливо завантажити об'єкти в контейнер, для подальшого зберігання;
- шифрування даних, які зберігаються – який рівень захисту має хмарне сховище на стороні клієнта та сервера;
- шифрування даних при передачі – який рівень захисту має хмарне сховище на рівні з'єднання;
- контроль доступу – чи є можливим надання обмеженого доступу користувачам для читання або завантаження нових об'єктів;
- управління версіями (Versioning) – чи є можливим використання для захисту об'єктів від ненавмисного перезапису та видалення управління версіями, зберігаючи кілька версій об'єкта в одному сегменті;
- реплікація даних – чи є автоматичною реплікація даних на різні регіони.

Сервіси забезпечення online - безпеки

У якості базової групи для розміщення блоків даних бази великих обсягів було обрано сервіси Amazon Web Services, Microsoft Azure та Google Cloud Platform.

Перш за все, за даними, котрі постачальники надають у відкритому доступі було проаналізовано механізми забезпечення захисту великих даних на означених online ресурсах.

На хмарних платформах існують різні сервіси забезпечення безпеки, ідентифікації та відповідності вимогам. Вони діляться на категорії:

- управління ідентифікацією;
- виявлення загроз та безперервний моніторинг;
- захист інфраструктури;
- захист даних.

Управління ідентифікацією – сервіси, які дозволяють безпечно керувати посвідченнями, ресурсами та дозволами у будь-якому масштабі. Хмарні платформи надають послуги з ідентифікації в додатках для робочих ресурсів і користувачів. Ці сервіси дозволяють швидко приступити до роботи і керувати доступом до робочих навантажень та додатків.

Виявлення загроз та безперервний моніторинг – виявлення загроз і безперервний моніторинг, хмарні платформи ідентифікують погрози, безперервно відстежуючи активність у мережі та поведінку в облікових записах хмарного середовища.

Захист інфраструктури – сервіси для захисту мережі та додатків дають можливість застосовувати докладні політики безпеки в точках контролю мережі в межах організації. Допомагають перевіряти та фільтрувати трафік, щоб уникнути несанкціонованого доступу на рівні кордонів хоста, мережі та програми.

Захист даних – сервіси, які допомагають захистити дані, акаунти і робочі навантаження від несанкціонованого доступу. Сервіси захисту даних забезпечують можливості для шифрування, управління ключами та виявлення загроз. Вони здійснюють постійний моніторинг, захист акаунтів і робочих навантажень.

Слід помітити, що обрані серівіси практично ідентичні у всіх аспектах, однак AWS відстає з точки зору конфіденційності даних через відсутність рішень щодо анонімізації даних. Microsoft Azure надала найнижчі показники за швідкістю прийому/передачі даних. Google Platform же відстає за відповідністю та нормативними вимогами, ускладнено надання користувачеві певного сервісу з резервного копіювання та геоблокування для захисту даних від втручання зі сторони. Узагальнений вигляд рішень хмарних платформ по кожній з категорій наведено у таблиці 1.

За вище описаними критеріями було проведено й практичну (загальну) оцінку перелічених об'єктних хмарних сховищ. Критеріальна оцінка показала, що сервіси майже ідентичні. Але на відміну від Amazon та Google Platform, Microsoft Azure не має автоматичного управління версіями. Користувачу потрібно вмикати його вручну, також він має менший відсоток по угоді про рівень обслуговування. Перевагою Google Cloud Platform є більша кількість доступних зон та регіонів.

Таблиця 1 – Рішення хмарних платформ за категоріями

Категорії	AWS	Azure	GCP
Відповідність та нормативні вимоги	CloudFront, KMS, data lifecycle manager	Front Door, Key Vault, Azure Backup, blob storage lifecycle	Cloud Key Management
Ідентифікація, автентифікація та авторизація	Cognito, SSO, client VPN, security token service, IAM, CloudWatch	AD SSO, multi-factor, IAM, AD token service, access control service	Identify Platform, Cloud Identity, VPN, Cloud Logging, IAM
Безпечна розробка, експлуатація та адміністрування	Bastion host, GuardDuty, WAF, CloudWatch, AlientVault USM.4	Bastion host, advanced threat protection, WAF, Azure Monitor, Key Vault	SSH from the browser, Security Command, Cloud Armor, Cloud Monitoring
Конфіденційність	Certificate manager, KMS	Key Vault, Dynamic data masking on SQL database	Certificate Authority Service, Cloud Data Loss Prevention
Безпечна архітектура	VPC, WAF, IoT Device Management, Glacier, KMS, CloudHSM	Virtual Network, Key Vault, HSM, Coldblob storage	VPN, Cloud Armor, Cloud IoT, Cloud Storage, Cloud HSM, Cloud Key Management

Безпека та відповідність у хмарі

Безпека та відповідність – це спільна відповідальність між хмарними платформами і користувачем.

Розмежування відповідальності між платформою та користувачем відбувається наступним чином.

Хмарні провайдери юридично беруть на себе відповідальність за контроль працездатності апаратного забезпечення та глобальної інфраструктури (регіони, доступні зони, прикордонне розмежування); компонентів програмного забезпечення (захист хмарної операційної системи, засобів виконання обчислень, створення та супроводу баз даних, інфраструктури збереження) та мережних комунікацій. Це відповідальність за безпеку хмари.

В той же час користувачі повинні зосередити свою увагу на керуванні доступом, конфігуруванні мережі та брандмауера, шифруванні даних, забезпечені процедур автентифікації, ідентифікації, авторизації та доступу до даних. Це відповідальність за безпеку в хмарі.

Ця спільна модель повинна допомогти та полегшити операційний тягар користувача, оскільки хмарні платформи керують та контролюють компоненти від операційної системи хоста та рівня віртуалізації до фізичної безпеки об'єктів, на яких працює сервіс.

Кількість даних, котрими повинен управляти користувач, також є достатньо великою. Й, з урахуванням наявності у будь-якого підприємства конфіденційної інформації, котру користувач не завжди довіряє стороннім провайдерам, автори роботи розширили методику розподілу застосуванням приватної хмари. Для обрання й побудови такої структури проведено додаткове дослідження.

Обрання та застосування приватної хмари

Окрім використання послуг хмарних платформ, таких як Google Cloud Platform, Microsoft Azure та Amazon Web Services вирішено використати приватну хмару, яка може бути застосована для зберігання конфіденційних та технологічних даних щодо, наприклад, розміщення блоків даних на хмарах, розроблено додатковий програмний модуль.

Оцінювалися рішення: Cozy, Tonido, OwnCloud, NextCloud, Seafile, Ressilio Sync. Найбільшу оцінку отримав Ressillio Sync, але за описом, рішення орієнтовано на синхронізацію файлів, а не для тривалого зберігання великих даних, тому для даної цілі обрано хмарне сховище NextCloud. Рішення має необхідний функціонал, оптимальну швидкість, зручний інтерфейс, захищає дані з використанням алгоритму AES-256 та виходячи з офіційної документації розробників, підходить для роботи з даними великих обсягів.

У запропонованій методиці, графічне відображення якої наведено на рис. 3 дані розділяються на менші блоки, які зберігатимуться на різних хмарних структурах один за другим. Номери блоків, імена файлів, міста їх розташування, ознаки цілісності та параметри доступу залишаються на приватній хмарі й їх захист та збереження цілком полягає на володаря інформації.

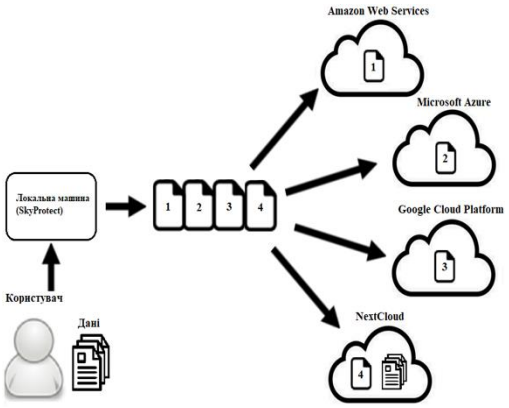

Рисунок 3 – Графічне відображення розподілу

Програмна реалізація методики

Для досягнення описаної методики було розроблено програмний додаток SkyProtect, основна мета якого - розподіл та опціонне шифрування й, за вимогою, реплікація даних великих обсягів серед об'єктних сховищ хмарних провайдерів (Amazon Web Services, Microsoft Azure, Google Cloud Platform) та на раніше обрану серед існуючих рішень шляхом проведення оцінки у заздалегідь підготовленому програмному модулі та самостійно розгорнуту приватну хмару NextCloud.

Програмний додаток написаний таким чином, щоб у майбутньому була можливість додавання додаткового приватного або об'єктного сховища хмарного провайдера.

Для початку роботи з програмним продуктом, необхідно створити акаунти на Amazon Web Services, Microsoft Azure та Google Cloud Platform. Після цього на кожній платформі необхідно отримати ключі доступу до API.

Для знаходження ключів Amazon s3, потрібно перейти до Консолі керування AWS, натиснути на службу S3, після цього у верхньому правому куті обрати ім'я свого облікового запису та перети до облікових даних безпеки. Після цього згенерувати свої ключі доступу та секретні ключі.

Для знаходження ключів Google Storage, потрібно перейти до Google API Console, потім до роздільника Google Cloud Storage. Обираємо API & Services, у розділі знаходимо підрозділ Credentials і там знаходимо ключі доступу.

Для знаходження ключів Windows Azure, потрібно перейти на портал Windows Azure. Спочатку потрібно створити новий проект сховища. Після вибору цього нового проекту внизу сторінки можна знайти керування ключами. У цьому випадку ключ доступу — це назва проекту сховища, а секретний ключ — первинний ключ в управлінні ключами.

Програмний додаток має два режими: основний та тестовий. При виборі основного режиму використовуються чотири різні хмари. Це вже згадані Amazon S3, Azure Blob Storage, Google Storage та NextCloud. Крім того, при тестовому режимі будуть використовуватися чотири папки в приватній хмарі NextCloud.

Використання різноманітних хмар вимагає, щоб програмний додаток SkyProtect мав справу з неоднорідністю інтерфейсів кожного хмарного сховища. Для реалізації драйверів до кожного об'єктного сховища хмарного провайдера або приватної хмари був написаний загальний інтерфейс CloudDriver.

Кожне хмарне сховище моделюється як пасивний об'єкт зберігання, який підтримує чотири операції:

- downloadData (завантажити дані з хмари в визначену директорію);
- uploadData (завантажити дані на хмару з визначеної директорії);
- deleteData (видалити окремий файл);
- deleteContainer (видалити DataUnit).

Для додавання нових файлів або завантаження існуючих необхідно обрати контейнер SkyProtectDataUnit (DataUnit). Контейнер повинен мати заголовний блок, в який включено: унікальну назву, номер версії (для підтримки завантаження додаткових версій файлів), дані для перевірки (хеш код усієї інформації всередині файлу), цифровий підпис (для додаткової перевірки цілісності файлу). Все вищезазначене буде зберігається в файлі metadata. Сам файл розміщуються в тому ж контейнері після заголовного блоку. При завантаженні на хмари, разом з файлами завантажуються метадані на кожну хмару. В разі, якщо метадані на одній хмарі пошкодяться, їх завжди можна буде зчитати з інших хмар. Для ініціалізації та перевірки цифрового підпису буде використовуватися алгоритм RSA з хеш функцією SHA-256. Зміст метаданих та контейнера DataUnit приведено на рисунку 4.

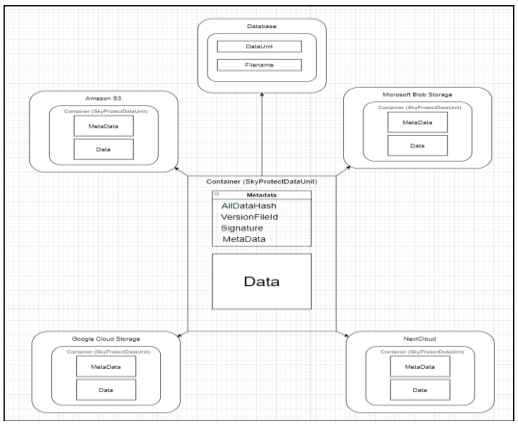

Рисунок 4 – Зміст контейнера SkyProtectDataUnit

Дані, що зберігаються в data unit, можуть мати довільний розмір, і цей розмір може бути різним для різних версій. Кожен data unit підтримує

звичайні операції зберігання об'єктів: створення (створення контейнеру і файлу метаданих з версією 0), знищення (видалення або видалення доступу до даних з версією 0), завантаження.

Окремим питанням забезпечення захисту великих обсягів є криптографічний захист. В якості опціонного шифрування реалізовано алгоритм шифрування AES на стороні користувача як найбільш відповідний для роботи з великими даними.

На Amazon S3, Microsoft Blob Storage, Google Cloud Storage та NextCloud підключаємо шифрування на стороні сервера, в кожній з хмар автоматично підключене шифрування AES-256, додатково на Microsoft Blob Storage можливе підключення другого рівня шифрування до даних – інфраструктурне шифрування.

Застосовані засоби захисту для приватної хмари перелічено в Таблиці 2.

Таблиця 2 – Засоби захисту приватної хмари

Категорія	Заходи
Управління ідентифікацією	Багатофакторна аутентифікація (MFA)
Виявлення загроз та безперервний моніторинг	Комплекс заходів згідно ДСТУ ISO/IEC 27000,
Захист інфраструктури	Політики інформаційної безпеки
Захист даних:	
- при збереженні	Шифрування AES-256, пакет Boxcriptor
- захист в процесі прийому/передачі	Протокол SSL/TLS
- в процесі обробки	Обробка шифрованих даних

Таким чином, перелік захисних заходів забезпечення безпеки приватної хмари відповідає вимогам міжнародного й вітчизняного законодавства у галузі інформаційних технологій.

Функціональність додатку демонструє діаграма варіантів використання, наведена на рис. 5, на якій видно, що система має одного актора (користувач), який може виконувати декілька дій: увійти в систему, вивести список команд на екран, вивести список існуючих файлів. Після вибору контейнера (SkyProtectDataUnit) користувач може завантажити файли або вивести обраний файл на екран. За складною структурою діаграми можна сказати, що програма має завершений вигляд, тобто користувач не буде обмежений функціональністю даного програмного додатку.

В процесі тестування в якості вхідних даних обрані бази даних розміром від 500 МБ до 20 ГБ Проведено тести з заміром часу шифрування та розшифрування, швидкості завантаження блоків даних на хмару та з хмари, з додатковою реплікацією та без неї. Контроль цілісності інформації виконувався при імітації атаки посередників. Контроль доступу проводився за спеціально складеною тестовою послідовністю з декілька сотень вірних та невірних спроб входів.

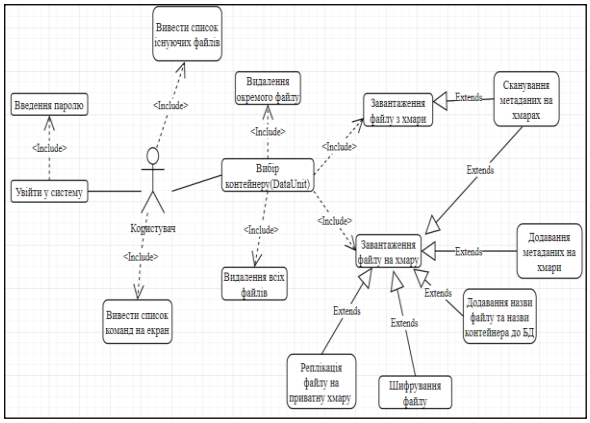

Рисунок 5 – Діаграма варіантів використання

За результатами тестування та дослідження сформульовано рекомендації.

Ресурси Amazon S3, NextCloud, Google Storage надали достатньо схожі результати й з урахуванням наявних інструментів захисту даних можуть бути застосовані для розміщення й обробки блоків даних будь я кого з обраних рівнів. Однак, застосування Amazon вимагає додаткових заходів з забезпечення конфіденційності даних. Azure Blob Storage показав більший час завантаження даних й цей ресурс й тому Azure пропонується застосовувати для розміщення не часто змінюваних типів даних, наприклад, довідників, специфікацій або таблиць постійних значень. Google Storage не має вказаних недоліків, але не пропонує засобів для проведення спеціалізованої обробки даних (наприклад, методами інтелектуального аналізу).

Висновки

В роботі досліджено стан питання забезпечення безпеки даних великих обсягів.

Проаналізовано механізми забезпечення захисту даних на online-ресурсах різного типу.

Виконано опис критеріїв вибору online-ресурсів різних моделей розгортання з точки зору гарантування відповідності критеріям безпеки;

Описано методику розміщення даних великих обсягів та додаток для розбивки, розподілу та реплікації даних на online-ресурсах

Наведено структуру та опис програмного забезпечення поділу, розподілу та опціонного шифрування та реплікації даних великих обсягів серед об'єктних сховищ різних моделей.

Зроблено ще один крок у напрямку дослідження можливостей забезпечення захисту технологій великих даних з застосуванням online-ресурсів різних моделей розгортання

Список літератури

1. Agarwal S., Gupta M., Sharma A. Big Data Privacy Issues Solutions. Proc. IEEE Int. Conf. Image Inf. Process. 2019. Vol. 2019-Novem, no. March, P. 225–228.
2. Федорко М.А., Маслова Н.О. Особливості захисту даних великих обсягів: наукові праці ДонНТУ. Серія "Інформатика, кібернетика. та обчислювальна техніка". 2018. No 1 (26), С.41-48.
3. The 2020 Data Attack Surface Report, Arcserve, 2 Dec 2020, URL: https://info.arcserve.com/en/the-2020-data-attack-surface-zeport#:~:text=Cybersecurity%20Ventures%20research%20shows%20the,world's%20data%20at%20that%20time.
4. Орлов В. Технология Google MapReduce и ее реализация с открытым исходным кодом Hadoop. URL: https://lvee.org/ru/reports/LVEE_2010_19.
5. Что такое Apache Hadoop в Azure HDInsight? URL: https://docs.microsoft.com/ru-ru/azure/hdinsight/hadoop/apache-hadoop-introduction.
6. Модуль Hadoop-AWS: интеграция с Amazon Web Services. URL: https://docs.arenadata.io/adh/administration/hdfs/AWS.html.
7. The Apache Software Foundation. Apache Hadoop, 28 July, 2020, URL: https://blogs.apache.org/hadoop/entry/ announce-apache-hadoop-3-3.
8. Маслова Н., Половинка О. Hadoop-рішення для захисту даних великих обсягів. Фізико-математичне моделювання та інформаційні технології, 2021, Вип. 33. С. 23-27, URL: http://www.fmmit.lviv.ua/index.php/fmmit/issue/view/32.
9. NBDPWG - Security and Privacy Subgroup, "NIST Special Publication 1500-4: Security and Privacy," NIST Big Data Interoperability Framework. 2019. Vol. 4.
10. Dev H., Sen T., Basak M., Ali M. E. An approach to protect the privacy of cloud data from data mining based attacks: Proceedings - 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, 2012. No. November, P. 1106–1115. doi: 10.1109/SC.Companion.2012.133.
11. Suwansrikham P., She K. Protection of big data privacy on multiple cloud providers by asymmetric security scheme: ACM International Conference Proceeding Series, 2019. No. November, P. 47–53. doi: 10.1145/3354153.3354156.
12. Rong C., Cheng H., Jaatun M. G. Securing big data in the Cloud by protected mapping over multiple providers: 2016 Digital Media Industry and Academic Forum, DMIAF 2016 - Proceedings, 2016. P. 166–171. doi: 10.1109/DMIAF.2016.7574925.
13. Viswanath G., Krishna P. V. Hybrid encryption framework for securing big data storage in multi-cloud environment: Evolutionary Intelligence. 2020. doi: 10.1007/s12065-020-00404-w.

References

1. Agarwal, S, Gupta, M, Sharma, A. (2019), "Big Data Privacy Issues Solutions", *Proc. IEEE Int. Conf. Image Inf. Process.*, vol. 2019-Novem, no. March, pp. 225–228.
2. Fedorko, M.A., .Maslova, N.O. (2018), "Features of data protection of large volumes" [Osoblyvosti zakhystu danykh velykykh obsyahiv], Naukovi pratsi DonNTU. Seriya "Informatyka, kibernetyka. ta obchyslyuval′na tekhnika" № 1 (26), pp.41-48.
3. The 2020 Data Attack Surface Report, Arcserve, 2 Dec 2020, available at: https://info.arcserve.com/en/the-2020-data-attack-surface-report#:~:text=Cybersecurity%20Ventures%20research%20shows%20the,world's%20data%20at%20that%20time.
4. Orlov, V. (2010), Google MapReduce technology and its implementation with open source Hadoop [Tekhnologiya Google MapReduce i yeye realizatsiya s otkrytym iskhodnym kodom Hadoop], available at: https://lvee.org/ru/reports/LVEE_2010_19 .
5. What is Apache Hadoop in Azure HDInsight [Chto takoye Apache Hadoop v Azure HDInsight], available at: https://docs.microsoft.com/ru-ru/azure/hdinsight/hadoop/apache-hadoop-introduction.
6. Hadoop-AWS module: integration with Amazon Web Services [Modul' Hadoop-AWS: integratsiya s Amazon Web Services], available at: https://docs.arenadata.io/adh/administration/hdfs/AWS.html.
7. The Apache Software Foundation. Apache Hadoop, 28 July, 2020, available at: https://blogs.apache.org/hadoop/entry/announce-apache-hadoop-3-3.
8. Maslova, N., Polovynka, O. (2021), *Hadoop-solutions for data protection of large volumes. Physical and mathematical modeling and information technologies* [*Hadoop-rishennya dlya zakhystu danykh velykykh obsyahiv. Fizyko-matematychne modelyuvannya ta informatsiyni tekhnolohiyi]*, Vol. 33, pp 23-27, available at: http://www.fmmit.lviv.ua/index.php/fmmit/issue/view/32.
9. NBDPWG - Security and Privacy Subgroup (2019), "NIST Special Publication 1500-4: Security and Privacy," NIST Big Data Interoperability Framework, Vol. 4.

10. Dev, H., Sen, T., Basak, M., Ali, M. E. (2012), An approach to protect the privacy of cloud data from data mining based attacks, *Proceedings 2012 SC Companion: High Performance Computing, Networking Storage and Analysis*. No. November, pp. 1106–1115. doi: 10.1109/SC.Companion.2012.133.
11. Suwansrikham, P., She, K. (2019), "Protection of big data privacy on multiple cloud providers by asymmetric security scheme," *ACM International Conference Proceeding Series*, No. November, pp. 47–53, doi: 10.1145/3354153.3354156.
12. Rong, C., Cheng, H., Jaatun, M. G. (2016), "Securing big data in the Cloud by protected mapping over multiple providers," *2016 Digital Media Industry and Academic Forum, DMIAF 2016 - Proceedings*, pp. 166–171 doi: 10.1109/DMIAF.2016.7574925.
13. Viswanath, G., Krishna, P. V. (2020), "Hybrid encryption framework for securing big data storage in multi-cloud environment," *Evolutionary Intelligence*. doi: 10.1007/s12065-020-00404-w.

Надійшла до редакції 03.11.2021

[1]N. MASLOVA, [2]A. NIKITENKO
Donetsk National Technical University, Pokrovsk, Ukraine
[1]nataliia.maslova@donntu.edu.ua
[2]andrii.nikitenko.kita@donntu.edu.ua

FEATURES OF PROTECTING BIG DATA ON ONLINE RESOURCES

Big data is one of the challenges of our time associated with the need to process ever-growing volumes of information. The trend when working with big data is the use of cloud or online resources. The most important aspect when working in a cloud environment is to ensure security - a computer, a network, information. The paper proposes a methodology and software tool for placing data in cloud environments. The basic condition for the placing Big data are the requirements for compliance with security criteria.

Keywords: Big data, information security, cloud environments, on-line resources.