# ХМАРНІ ТЕХНОЛОГІЇ У СПОРТІ: ЗБЕРІГАННЯ, ОБРОБКА ТА БЕЗПЕКА ДАНИХ

Компанієць М.Г.

Харківський національний університет радіоелектроніки,
Україна.

E-mail: mykhailo.kompaniiets@nure.ua

**Abstract**

The integration of digital technologies has fundamentally reshaped the sports industry, with cloud computing emerging as a central paradigm for managing and processing vast amounts of athlete data. This shift is driven by the economic efficiency of cloud platforms, the proliferation of Internet of Things (IoT) sensors, and the scalability of machine learning (ML) algorithms, which create new lucrative business models based on data analytics. However, this reliance on cloud infrastructure for storing sensitive physiological, medical, and strategic data introduces significant security, privacy, and legal dangers. This paper provides a comprehensive analysis of the advantages and dangers of using cloud platforms for sports analytics, focusing on the storage, processing, and security of training telemetry, medical data, and video analysis. The methodology is based on a systematic review of current research analyzing cloud architectures, specific service platforms, security threats, and legal frameworks. The findings indicate that the primary advantages of cloud adoption are the scalability and economic efficiency of automated machine learning (AutoML) platforms, such as AWS SageMaker, GCP Vertex AI, and MS Azure, which provide high-accuracy analytics (e.g., 99.6% accuracy for XGBoost on SageMaker) at a reduced cost and development time. These platforms enable global, remote access to performance insights, democratizing advanced analytics for both professional and amateur athletes. Conversely, these platforms present significant dangers. Analysis reveals inconsistent model accuracy (ranging from 84.2% to 99.6%), with some providers utilizing proprietary algorithms. The most critical danger identified is Data Breaches, ranked as the top threat by the Cloud Security Alliance. This risk is exacerbated by vulnerabilities such as Insecure APIs and Account Hijacking.

Інтеграція цифрових технологій в управління спортивними ресурсами спричинила фундаментальні зміни, революціонізувавши підходи до операційної діяльності спортивних організацій та аналізу показників атлетів. Хмарні обчислення стали ключовим елементом цієї трансформації. Вони надають необхідну гнучкість, масштабованість, економічну ефективність та обчислювальну потужність для обробки великих наборів даних, усуваючи потребу в утриманні дорогої та складної локальної IT-інфраструктури. Одночасний розвиток алгоритмів машинного навчання (ML), зокрема платформ автоматизованого ML, створює продуктивні умови для масштабування. Дослідження показують, що хмарні рішення є більш економічними та значно скорочують час на розробку аналітичних моделей порівняно з традиційними підходами. Ця технологічна доступність обслуговує широкий спектр користувачів - від професійних спортивних ліг до аматорських та рекреаційних організацій. Зростаючий попит на аналітику, зумовлений збільшенням кількості атлетів, у поєднанні зі здешевленням обробки даних, формує нові прибуткові бізнес-моделі. Таким чином, виникає позитивний цикл: підвищення якості та доступності аналітичних послуг стимулює залучення ширших верств населення до спорту, що, у свою чергу, розширює ринок для подальшого розвитку цих технологій.

## Зберігання та обробка телеметрії спортсменів

Збір даних про спортсменів охоплює широкий спектр телеметрії, що вимагає гнучких рішень для зберігання та потужних інструментів для обробки. Хмарні платформи агрегують дані з різноманітних джерел для досягнення двох головних цілей: аналізу продуктивності та прогнозування травм. Ключовим викликом при цьому є не лише різноманітність (гетерогенність) даних, але й їхній експоненційно зростаючий обсяг та швидкість надходження. Сучасні IoT-сенсори та тілесні мережі

(BAN) генерують безперервні потоки телеметрії з високою частотою дискретизації. Наприклад, дані тривимірного прискорення (jerk) або показники з гіроскопів можуть збиратися сотні разів на секунду (Гц), створюючи терабайти сирих даних за одне тренування або змагання. Саме ця необхідність агрегувати, зберігати та обробляти такі масиви даних змушує спортивні організації відмовлятися від локальних серверів на користь масштабованої хмарної інфраструктури. Основним типом даних, що збирається є фізіологічна телеметрія. Носимі пристрої та датчики тілесної мережі збирають показники в реальному часі. До них належать частота серцевих скорочень, GPS-координати та тривимірне прискорення, яке використовується для аналізу постурального балансу та ризику травм.

Біомеханічні дані є другим по поширеності видом даних, що використовуються для аналітики. З фітнес-зображень та відео (до 6 мільйонів зразків в одному дослідженні) за допомогою таких бібліотек, як Mediapipe, витягуються координати суглобів. Медичні та суб'єктивні дані включають історію попередніх травм, наприклад, струсів мозку, а також суб'єктивне сприйняття стану атлетом, наприклад, за Спортивним фітнес-індикатором.

Сама обробка цієї акумульованої телеметрії для досягнення аналітичних цілей відбувається у двох фундаментально різних режимах. По-перше, це обробка в реальному часі, яка аналізує потокові дані безпосередньо під час їх надходження. Це дозволяє тренерам миттєво отримувати сповіщення, наприклад, про критичне падіння рівня витривалості гравця або про біомеханічні патерни руху, що свідчать про високий ризик травми в момент аналізу. По-друге, це пакетна обробка, яка застосовується до великих історичних масивів даних у стані спокою. Саме в цьому режимі відбувається навчання складних моделей машинного навчання, наприклад, ретроспективний аналіз мільйонів фітнес-зображень для створення класифікатора поз або аналіз GPS-даних за весь сезон для виявлення довгострокових тенденцій у продуктивності.

Основними цілями обробки даних є: аналіз продуктивності - відстеження показників для аналізу продуктивності гравця та оптимізації планів тренувань; діагностика та прогнозування - використання даних для діагностики спортивних травм та передбачення ймовірності майбутніх травм; класифікація - автоматичне розпізнавання та класифікація фітнес-поз на основі антропометричних даних.

Архітектура зберігання та сервіси обробки

Ефективна реалізація спортивної аналітики вимагає побудови багаторівневої хмарної архітектури, яка забезпечує безшовний потік даних від сенсора до аналітичної моделі. Фундаментальним рівнем цієї системи є периферійний рівень (Edge Layer), де тілесні мережі датчиків та IoT-пристрої здійснюють первинний збір телеметрії. Для задач відеоаналітики цей рівень використовує спеціалізовані сервіси, такі як Amazon Kinesis Video Streams (KVS), які забезпечують надійну потокову передачу відеоданих із тренувальних майданчиків у хмару для подальшої обробки.

На рівні отримання та оркестрації (Ingestion & Orchestration) дані потрапляють у хмарне середовище через захищені шлюзи, такі як AWS IoT Core, що відповідає за авторизацію пристроїв та управління їх станом. Як зазначається в технічних рекомендаціях AWS, сучасні архітектури все частіше використовують безсерверну модель) для оркестрації процесів. Сервіси на кшталт AWS Lambda та AWS Step Functions автоматично запускають процеси обробки (наприклад, нарізку відео або вилучення метаданих) у відповідь на нові надходження даних, забезпечуючи операційну ефективність та масштабованість.

Центральним елементом системи є рівень даних, реалізований за концепцією «озера даних» (Data Lake) на базі об'єктних сховищ, таких як Amazon S3. Сюди стікаються як «сирі» відео- та сенсорні дані, так і результати їх попередньої обробки. Для швидкого пошуку та індексації результатів інференсу (наприклад, знайдених подій у матчі) архітектура часто доповнюється пошуковими рушіями, такими як Amazon OpenSearch Service, що дозволяє аналітикам миттєво знаходити потрібні фрагменти. Паралельно, часові ряди фізіологічних показників спрямовуються у спеціалізовані бази даних (наприклад, Amazon Timestream), оптимізовані для швидких запитів.

Вершиною архітектури є рівень машинного навчання (ML Layer). На цьому етапі з неструктурованих даних витягуються ключові ознаки. Наприклад, з фітнес-зображень за допомогою бібліотеки Mediapipe витягуються координати суглобів (x, y, z). Ці структуровані дані подаються на вхід платформ AutoML, таких як AWS SageMaker, GCP Vertex AI або MS Azure. Емпіричні дослідження архітектур показують, що вибір платформи є критичним: наприклад, алгоритм XGBoost, розгорнутий на AWS SageMaker, продемонстрував точність 99.6% при класифікації вправ, тоді як аналогічні рішення на інших

платформах показали результат лише 84.2%. Це підкреслює необхідність ретельного вибору ML-сервісів залежно від специфіки задачі.

## Аналіз переваг та недоліків хмарних платформ

Ключовою перевагою хмарних технологій є суттєве зниження бар'єру входу для широкого кола користувачів, від професійних атлетів до аматорів. Хмарні обчислення пропонують гнучкі, масштабовані та економічно ефективні рішення, що усуває для організацій та індивідуальних користувачів потребу в інвестуванні у складну та дорогу локальну IT-інфраструктуру для зберігання й обробки даних. Це надає більш економічні, зручні та полегшені сервіси. Зокрема, використання платформ автоматизованого машинного навчання на хмарних сервісах, як-от AWS, GCP та Azure, є більш економічним та скорочує час на розробку моделей порівняно з традиційними методами ручної розробки алгоритмів.

Ця економічна доступність доповнюється перевагою глобального безперешкодного доступу до даних та аналітичних інтерфейсів. Архітектура хмарних обчислень за своєю суттю передбачає, що обчислювальні ресурси доступні віддалено користувачами через Інтернет. Для спортивних організацій це трансформується у можливість отримувати доступ до критично важливих даних з будь-якого місця. Тренери, аналітики та медичний персонал можуть безперешкодно обмінюватися даними про продуктивність гравців, записами ігор та звітами про стан здоров'я, забезпечуючи прийняття рішень у реальному часі незалежно від фізичного розташування. Водночас відбувається стрімкий розвиток самих алгоритмів обробки даних, що є фундаментом сучасної спортивної аналітики.

Хмарні платформи надають величезні обчислювальні потужності по запиту, необхідні для застосування складних алгоритмів та методів машинного навчання і штучного інтелекту. Дослідження демонструють використання хмарних систем з підтримкою глибокого навчання, зокрема глибоких згорткових нейронних мереж, для вилучення біомеханічних ознак та діагностики травм. Це призводить до прямого підвищення точності аналітики. Наприклад, порівняльний аналіз платформ AutoML для аналізу антропометрії вправ показав, що алгоритм XGBoost, реалізований на AWS SageMaker, досяг видатної точності 99.6% у класифікації. Збільшення користувацької бази та обсягів даних стимулює подальше розширення функціоналу, що стає можливим завдяки масштабованості хмарної інфраструктури. Хмарні сервіси дозволяють організаціям динамічно коригувати свої обчислювальні ресурси для обробки періодів високого трафіку.

Ця здатність до миттєвого масштабування дозволяє провайдерам відкривати нові потенціали та диверсифікувати джерела доходу, інвестуючи у розробку нових аналітичних інструментів, тим самим постійно розширюючи функціонал, доступний для кінцевих користувачів.

Попри значні переваги, використання хмарних платформ у спорті пов'язане з суттєвими ризиками, що охоплюють як технічні, так і безпекові аспекти. Надійність аналітики не є уніфікованою. Порівняльні дослідження платформ AutoML демонструють значні розбіжності у продуктивності: точність класифікації вправ може варіюватися від виняткових 99.6% (наприклад, XGBoost на AWS SageMaker) до значно нижчих 84.2% (LightGBM на MS Azure). Це доводить неможливість сліпо покладатися на будь-яке хмарне рішення. Більше того, для користувача ці сервіси часто функціонують як «чорна скринька». Деякі провайдери, наприклад GCP Vertex AI, можуть використовувати пропрієтарні алгоритми, що робить неможливим для кінцевого користувача аудит чи повне розуміння аналітичної моделі. Найбільш серйозною небезпекою є ризик витоку даних (data breaches).

Альянс хмарної безпеки (CSA) визначає це як загрозу номер один. Оскільки хмарні обчислення передбачають зберігання конфіденційної інформації на віддалених серверах, дані стають вразливими у випадку, якщо ці сервери скомпрометовані. Це не теоретичний ризик; реальні інциденти, як-от масштабний витік даних Capital One, спричинений діями колишнього інженера Amazon, та витік даних Facebook, що стався через незахищену конфігурацію хмарного сховища, ілюструють практичні небезпеки. Ці загрози, поряд із перехопленням облікових записів, атаками інсайдерів та атаками типу Man-in-the-Middle, наражають на небезпеку біометричні та медичні дані атлетів.

Додатковою проблемою є обмежена функціональність стандартних сервісів. Хоча платформи AutoML спрощують інженерний процес, вони можуть не задовольняти потреби просунутих аналітиків. Створення глибоко кастомізованих моделей вимагає просунутих знань у різних сферах, таких як мови кодування та предиктивний аналіз. Більшість існуючих інструментів зосереджені на загальних показниках, таких як продуктивність атлета, залишаючи значну прогалину в складніших інтегрованих

аналізах, наприклад, управлінні фінансовими ресурсами. Це призводить до того, що для ефективного використання даних потрібна спеціалізована експертиза, яка часто відсутня в традиційному спортивному середовищі, що підвищує ризик неправильної інтерпретації складних аналітичних звітів.

Однією з найсерйозніших, але часто ігнорованих загроз, є проблема прив'язки до постачальника (Vendor Lock-in). Використання пропрієтарних API та форматів даних створює ситуацію, коли міграція до іншого провайдера стає економічно невигідною або технічно неможливою через високу вартість адаптації. Це робить спортивні організації залежними від цінової політики та технологічної дорожньої карти одного монополіста.

Іншим критичним недоліком є мережева затримка (Latency). Хоча хмари ідеальні для пакетної обробки, передача даних на віддалені сервери та отримання відповіді може займати сотні мілісекунд, що є неприпустимим для тактичних рішень у реальному часі. Дослідження показують, що хмарна модель може програвати в швидкодії периферійним обчисленням, які здатні зменшити затримку до 80% завдяки обробці даних безпосередньо на місці події.

Фінансова непередбачуваність також стає новим викликом. Хоча початковий вхід є дешевим, складні моделі ціноутворення провайдерів часто призводять до неочікуваних витрат, особливо за вихідний трафік (egress fees) або «холодний» запуск безсерверних функцій. Дослідження динаміки витрат вказують, що для високомасштабних, постійно працюючих аналітичних систем хмарна модель може виявитися менш ефективною за фіксовану інфраструктуру, якщо не застосовувати суворі практики FinOps.

Фундаментальною вразливістю архітектури залишається залежність від стабільності мережевого з'єднання. Хоча сучасні протоколи передачі даних (наприклад, MQTT) підтримують локальну буферизацію на сенсорах, тривалі збої зв'язку створюють ризик переповнення пам'яті IoT-пристроїв та незворотної втрати телеметрії. Крім того, при відновленні з'єднання виникає проблема узгодженості даних (Data Consistency): необхідність синхронізувати локальні кеші з хмарою може призвести до конфліктів версій та затримок у доступності актуальної аналітики для тренерського штабу.

## Виклики та комплексність надання хмарних аналітичних послуг

Переваги хмарних технологій для кінцевого користувача створюють значні та багатогранні виклики для компаній-провайдерів (Cloud Service Providers, CSPs), які розробляють, підтримують та надають ці послуги. Ці виклики виходять за межі суто технічної реалізації алгоритмів і охоплюють складні сфери юридичного регулювання, управління безпекою та політики взаємодії з користувачами.

Найбільш вагомим викликом для провайдерів є адаптація до суворого юридичного регулювання, особливо в контексті обробки медичних та біометричних даних. Ключовим регуляторним актом є Загальний регламент ЄС про захист даних (GDPR). Його впровадження докорінно змінило ландшафт відповідальності для CSPs.

По-перше, GDPR має екстериторіальну застосовність (Стаття 3). Це означає, що будь-який провайдер, незалежно від його фізичного розташування (наприклад, американські компанії, як-от Google чи Amazon), юридично зобов'язаний дотримуватися норм GDPR, якщо він обробляє персональні дані суб'єктів, що перебувають в ЄС. Це особливо актуально для спортивних сервісів, що мають глобальну аудиторію. Регламент чітко кваліфікує генетичні дані та дані про здоров'я як спеціальні категорії, що вимагають посиленого захисту.

По-друге, GDPR вводить нову парадигму спільної відповідальності. Регламент чітко розмежовує ролі: контролер даних (організація, що визначає цілі обробки, наприклад, спортивний клуб) та обробник даних (провайдер CSP, що надає інфраструктуру). Якщо раніше відповідальність провайдера була мінімальною, то тепер, у випадку витоку даних, він несе спільну відповідальність разом із контролером. Це змушує CSPs фундаментально переглядати свої архітектури та контракти, впроваджуючи нові умови надання послуг (Terms of Service) та кодекси поведінки (Code of Conducts) для управління цими ризиками.

Правові норми безпосередньо впливають на політику взаємодії з користувачем. GDPR вимагає від компаній отримання чіткої позитивної дії або непасивної згоди від користувача на обробку даних. Це означає, що провайдери більше не можуть покладатися на попередньо встановлені прапорці у формах реєстрації.

Регламент надає користувачам право на забуття (right to be forgotten), тобто право вимагати повного видалення своїх даних. Це створює колосальний технічний виклик для провайдерів. Реалізація повного видалення даних є нетривіальною задачею в розподілених хмарних архітектурах, де дані можуть бути репліковані та фрагментовані у різних дата-центрах. На додаток до юридичних викликів, провайдери стикаються зі складною технічною задачею захисту інфраструктури. Загрози, такі як витоки даних, перехоплення облікових записів та зловмисні інсайдери, визначені як головні ризики для хмарних середовищ.

У сучасному мультихмарному середовищі, де дані можуть оброблятися у кількох провайдерів (наприклад, AWS та GCP), поверхня атаки розширюється. Компанії повинні захищати небезпечні API (Insecure APIs), що використовуються для зв'язку між цими хмарами. Стандартні системи виявлення вторгнень (IDS) часто виявляються неефективними, оскільки вони не враховують повний життєвий цикл обробки файлів, що ускладнює виявлення атак.

Ця комплексність змушує компанії розробляти інноваційні та дорогі методи захисту. Наприклад, дослідження пропонують використання майнінгу процесів (process mining) для аналізу логів. Цей підхід дозволяє відстежувати повний життєвий цикл файлу (створення, читання, переміщення між хмарами, видалення) та виявляти аномальні шляхи виконання або несанкціоновані операції, які традиційні IDS пропустили б. Таким чином, розробка та підтримка надійних алгоритмів аналітики нерозривно пов'язані з розробкою не менш складних алгоритмів безпеки.

## Висновки

Проведений аналіз демонструє, що хмарні технології стали невід'ємною рушійною силою цифрової трансформації у спортивній індустрії. Вони докорінно змінили підходи до зберігання та обробки великих масивів даних, надавши інструменти для глибокого аналізу продуктивності атлетів та прогнозування травм. Завдяки агрегації різноманітної телеметрії – від фізіологічних показників з носимих IoT-пристроїв до біомеханічних координат суглобів, отриманих з відео – хмарні платформи, такі як AWS SageMaker, GCP Vertex AI та MS Azure, уможливили застосування складних алгоритмів машинного навчання, як-от CNN та XGBoost, для високоточної діагностики.

Ключові переваги такого підходу полягають у масштабованості, економічній ефективності та глобальній доступності даних. Це фактично демократизувало доступ до потужної аналітики, раніше доступної лише професійним лігам, зробивши її доступною для аматорів та невеликих організацій.

Однак ці переваги нерозривно пов'язані зі значними небезпеками. По-перше, для користувача існує ризик неузгодженої точності та непрозорості алгоритмів, що функціонують як "чорна скринька". По-друге, що є більш критичним, зберігання конфіденційних медичних та біометричних даних на віддалених серверах створює постійну загрозу витоку даних. Реальні інциденти, як-от витоки даних Capital One та Facebook, підтверджують, що навіть провідні провайдери вразливі до атак або людських помилок.

Для самих компаній-провайдерів ці ризики трансформуються у комплексні технічні та юридичні виклики. Найвагомішим із них є Загальний регламент ЄС про захист даних (GDPR). Введення принципу спільної відповідальності (joint control) та вимоги непасивної згоди кардинально підвищують рівень відповідальності провайдерів.

Це, у свою чергу, стимулює розробку нових, більш складних систем безпеки. Стандартні засоби виявлення вторгнень (IDS) виявляються неефективними, оскільки вони не здатні відстежити повний життєвий цикл обробки файлів у складних мультихмарних середовищах, що робить їх вразливими до інсайдерських атак. Як наслідок, провайдери змушені інвестувати у дорогі технології, таких як майнінг процесів для аналізу аномалій у логах та забезпечення відповідності вимогам GDPR.

## Література

1. L. Xiao, Y. Cao, Y. Gai, J. Liu, P. Zhong, and M. M. Moghimi. Review on the application of cloud computing in the sports industry. Journal of Cloud Computing, vol. 12, no. 152, 2023.

2. X. Wu, J. Zhou, M. Zheng, S. Chen, D. Wang, J. Anajemba, G. Zhang, M. Abdelhaq, R. Alsaqour, and M. Uddin. Cloud-based deep learning-assisted system for diagnosis of sports injuries. Journal of Cloud Computing, vol. 11, no. 82, 2022.

3. A. Alotaibia, M. AlZaina, M. Masuda, and N. Jhanjhib. A Comprehensive Survey on Security Threats and Countermeasures of Cloud Computing Environment. Turkish Journal of Computer and Mathematics Education, vol. 12, no.9, 2021.

4. X. Zhang, L. Cui, W. Shen, J. Zeng, L. Du, H. He, and L. Cheng. File processing security detection in multi-cloud environments: a process mining approach. Journal of Cloud Computing, vol. 12, no. 100, 2023

5. B. Russo, L. Valle, G. Bonzagni, D. Locatello, M. Pancaldi, and D. Tosi. Cloud Computing and the new EU General Data Protection Regulation. IEEE Cloud Computing, vol. 5, p. 58-68, 2018.

6. M. Ahmed and M. Hossain. Cloud Computing and Security Issues in the Cloud. International Journal of Network Security & Its Applications, vol. 6, no. 1, 2014.

7. R. Narendra, S. Tadapaneni, M. Sabri. Cloud Computing Security Challenges. SSRN Electronic Journal, vol. 7, no. 6, 2020.

8. M. Romano. Digital Transformation in Sports Resource Management: Challenges and Opportunities. International Journal of Artificial Intelligence, Data Science, and Machine Learning, vol. 1, no. 3, 2020.

9. Guidance for Video Analysis as a Service on AWS. https://aws.amazon.com/solutions/guidance/video-analysis-as-a-service-on-aws

10. W. Choi, T. Choi, S. Heo. A Comparative Study of Automated Machine Learning Platforms for Exercise Anthropometry-Based Typology Analysis: Performance Evaluation of AWS SageMaker, GCP VertexAI, and MS Azure, 2023.

11. J. Opara-Martins, R. Sahandi, and F. Tian. Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. Journal of Cloud Computing, vol. 5, no. 4, 2016.

12. M. Hamza, M. Akbar, R. Capilla. Understanding Cost Dynamics of Serverless Computing: An Empirical Study, arXiv:2311.13242, 2024.

13. M. Satyanarayanan. The emergence of edge computing. Computer, vol. 50, no. 1, pp. 30-39, 2017.