

JEL C69; I28; I29; L86

Гужва Володимир

к.е.н., доцент,

Київський національний

економічний університет

ім. Вадима Гетьмана, Україна

ORCID: 0000-0002-0379-1480

e-mail: volodymyr.huzhva@kneu.ua

Volodymyr Huzhva

PhD, Associate Professor

Kyiv National Economic University

named after Vadym Hetman, Ukraine

УПРАВЛІННЯ ДАНИМИ В ЦИФРОВІЙ ТРАНСФОРМАЦІЇ АКАДЕМІЧНИХ УСТАНОВ: ВІД БАЗ ДО ФАБРИК ДАНИХ

DATA GOVERNANCE IN THE DIGITAL TRANSFORMATION OF ACADEMIC INSTITUTIONS: FROM DATABASES TO DATA FACTORIES

Анотація. Робота присвячена еволюції технологій управління даними в процесі здійснення цифрової трансформації в академічних установах – від баз даних до фабрик даних. Наведено аналіз їх застосування в академічних установах.

Ключові слова: академічна установа, цифрова трансформація, база даних сховище даних, озеро даних, фабрика даних.

Abstract. The work is devoted to the evolution of data management technologies in the process of digital transformation in academic institutions — from databases to data factories. An analysis of their application in academic institutions is presented.

Keywords: academic institution, digital transformation, database, data warehouse, data lake, data factory.

Загальноновизнаним трендом сучасного розвитку соціально-економічних систем в цілому і академічних установ зокрема є *цифрова трансформація* (ЦТ) [1]. Для успішної реалізації ЦТ у закладах освіти не можна обмежуватися лише одним оновленням технічної інфраструктури — метою повинна стати перебудова способів навчання, досліджень, адміністрування та взаємодії між учасниками освітнього процесу. Для досягнення зазначених результатів ЦТ потрібно використовувати різноманітні технологічні драйвери, до переліку яких слід віднести: *хмарні технології, штучний інтелект та машинне навчання, інтернет речей, блокчейн-технології, технології віртуальної реальності, технології управління даними* тощо.

Управління даними відіграє важливу роль в академічних установах, забезпечуючи цілісність, доступність і безпеку даних, що є основою для досліджень, навчання та розвитку [2]. До переліку факторів, що покликані акцентувати особливу увагу на питаннях управління даними, слід віднести: 1) *обсяги та різноманітність даних*; 2) *інтеграція різнотипних даних*; 3) *академічна прозорість та репутація*; 4) *прийняття рішень на основі даних*; 5) *міжнародна конкуренція та співпраця*.

Еволюцію форм збереження та управління даними в академічних установах коротко можна представити таким логічним ланцюгом: *паперові архіви – лінійні файли – бази даних – сховища даних – озера даних – фабрики даних*.

Бази даних (БД) виступають фундаментальним інструментом для збереження, обробки та управління великим обсягом інформації в академічному середовищі. Найбільш поширеним типом баз даних у академічних установах є *реляційні бази даних* (Relational Database — RDB), призначені для зберігання та обробки структурованих даних. Це можуть бути студентські записи, розклади, результати навчання, адміністративні дані (інформація про співробітників, фінансові операції, управлінські звіти). З ростом обсягів неструктурованих даних академічні установи почали впроваджувати *нереляційні бази даних* (Not Only SQL — NoSQL). Вони призначені для більш гнучкої роботи з даними. Бази даних є основою для багатьох систем управління навчанням (LMS), таких як Moodle, Blackboard, Canvas, які використовують реляційні або NoSQL бази для зберігання інформації про студентів, навчальні матеріали, оцінки та інші освітні дані. LMS дозволяє зберігати та керувати інформацією про курси, інтерактивні матеріали, завдання, результати іспитів та інші навчальні ресурси. Слід зазначити те, що мають місце і проблеми та виклики використання традиційних баз даних в академічних установах: 1) *масштабованість* — реляційні бази даних мають обмеження щодо масштабованості, особливо при роботі з великими обсягами неструктурованих даних; 2) *гнучкість* — реляційні бази даних не забезпечують достатньої гнучкості для роботи з неструктурованими даними або даними, які швидко змінюються; 3) *управління великими даними* — збільшення обсягу даних у навчальних процесах і дослідженнях потребує більш складних систем управління великими даними (Big Data).

З розвитком технологій управління даними в академічних установах і бізнесі зазнало значних змін — від традиційних баз даних до сучасних концепцій — Data Warehouses (сховищ даних), Data Lakes (озер даних) та Data Factory (фабрик даних). Кожен етап еволюції сприяв підвищенню ефективності роботи з інформацією.

Data Warehouse (DWH) — це централізоване сховище, що дозволяє зберігати, інтегрувати і аналізувати великі обсяги структурованих даних з різних джерел і призначене для підтримки процесів прийняття рішень і аналітики у великих організаціях. DWH забезпечують інтеграцію усіх операційних та наукових даних в установі у єдину систему, що дозволяє адміністратору або дослідникам отримати повну картину щодо навчальних процесів, результатів студентів, ефективності досліджень і управлінських процесів тощо [3]. Разом з тим слід навести і ряд основних викликів при використанні сховищ даних: 1) *складність налаштування* — сховища даних можуть бути складними у налаштуванні та потребувати значних ресурсів для інтеграції різних джерел даних; 2) *підготовка даних* — одним з головних обмежень сховищ є необхідність попередньої обробки та структурування даних перед їх завантаженням, що може вимагати додаткового часу і зусиль.

Концепція *озер даних* (Data Lake) з'явилися у відповідь на обмеження традиційних сховищ даних [4]. *Озеро даних* — це гнучке сховище, яке дозволяє зберігати необроблені дані з будь-якого джерела у будь-якому форматі. Це можуть бути структуровані дані, неструктуровані дані (зображення, відео, документи) або напівструктуровані дані (JSON, XML тощо). До переліку основних характеристики озер даних слід віднести: 1) *гнучкість у зберіганні* — дозволяють зберігати всі типи даних незалежно від їхньої структури або форми;

2) *збереження необроблених даних* — дані зберігаються у своєму початковому вигляді і можуть бути оброблені та використані пізніше за потреби для аналізу; 3) *масштабованість* — озера даних легко масштабуються, що дозволяє зберігати величезні обсяги інформації та ефективно управляти ними. Серед викликів, пов'язаних з використанням озер даних, слід виділити наступні: а) *проблеми з якістю даних* — оскільки дані не обробляються під час збереження, є ризик зберігання неякісних або дублікатів даних, що може ускладнити подальший аналіз; б) *потреба в складних аналітичних інструментах* — для ефективного використання даних з озера потрібно застосовувати складні інструменти для пошуку, обробки та аналізу, що може вимагати значних ресурсів.

Наступним етапом у розвитку технологій збереження та управління даними є *фабрики даних* (Data Factory). Фабрики даних є хмарними або гібридними платформами. Вони орієнтовані на виконання основних функцій, які дозволяють автоматизувати управління даними: а) *ETL (Extract, Transform, Load)* — процес витягування даних з різних джерел, їх трансформації відповідно до визначених бізнес-правил та завантаження; б) *інтеграція різних джерел даних* — фабрики даних дозволяють об'єднувати дані з різних джерел, таких як хмарні сервіси, локальні бази даних, системи бізнес-аналітики або системи управління навчанням (LMS); в) *оркестрація процесів* — фабрики даних дозволяють налаштовувати та автоматизувати процеси обробки даних за допомогою візуальних інтерфейсів та програмованих завдань; г) *моніторинг та керування* — платформи для фабрик даних пропонують інструменти для моніторингу стану процесів, виявлення помилок і аналізу ефективності роботи.

На ринку існує кілька потужних платформ фабрик даних, серед яких виділяються: 1) *Azure Data Factory (ADF)* — хмарне рішення від Microsoft для автоматизації ETL-процесів та створення конвеєрів даних для обробки даних з можливістю використання штучного інтелекту та машинного навчання для додаткового аналізу; 2) *Amazon Web Services (AWS) Glue* — платформа для автоматизованого витягування, трансформації та завантаження даних, що інтегрується з іншими хмарними сервісами AWS; 3) *Google Cloud Dataflow* — інструмент для обробки потоків даних (dataflow) і пакетної обробки в реальному часі на базі Google Cloud Platform.

Академічні установи все частіше використовують Data Factories для управління своїми даними через кілька ключових причин: 1) *необхідність інтеграції різних джерел даних* — академічні установи збирають дані з багатьох джерел, таких як системи управління навчанням (LMS), наукові репозитарії, дослідницькі бази даних, адміністративні системи. Фабрики даних дозволяють ефективно інтегрувати всі ці джерела у єдиний аналітичний простір; 2) *потреба в безперервній автоматизації обробки* — фабрики даних автоматизують процеси обробки великих обсягів інформації, що важливо для ефективного управління студентськими записами, науковими даними, результатами досліджень та аналітикою діяльності установи; 3) *підтримка великих даних (Big Data)* — багато академічних проєктів, особливо у сфері природничих наук, генерують великі обсяги даних. Фабрики даних дозволяють ефективно управляти та аналізувати ці дані в реальному часі, використовуючи можливості хмарних обчислень і високо-

продуктивних обробників даних; 4) *забезпечення гнучкості для аналізу* — фабрики даних підтримують різноманітні формати даних, що дає змогу академічним дослідникам працювати з різнотипними даними, включаючи структуровані (таблиці, бази даних) та неструктуровані (тексти, зображення, відео) дані.

Потрібно також назвати і можливі проблеми при впровадженні фабрик даних в академічних установах: а) *складність налаштування*; б) *управління якістю даних*; в) *вартість*.

Таким чином, ефективне управління даними в умовах цифрової трансформації є одним із ключових факторів успіху сучасних академічних установ. Відмова від традиційних підходів і перехід до нових технологій, таких як фабрики даних, дозволяє не лише оптимізувати процеси, але й сприяти науковому розвитку, інноваціям та підвищенню академічної конкурентоспроможності на глобальному рівні.

Література

1. Zabalawi, I., Kordahji, H., Aftimos, S. Digital Transformation in Digital Transformation in Universities: Strategic Universities: Strategic Framework, Implementation Framework, Implementation Tools, and Leadership. URL: Tools, and Leadership. URL: https://doi.org/10.1007/978-3-031-70779-7_8
2. Michaela Morgan. Achieving Data Governance in Higher Education. URL: Governance in Higher Education. URL: <https://panteon.io/blog/data-governance-in-higher-education>
3. An overview of the role of data warehouses in higher education IT systems. URL: <https://lumendata.com/blog/an-overview-of-the-role-of-data-warehouses-in-higher-education-it-systems/>
4. Diana Martinez-Mosquera, Verónica Beltrán; Diego Riofrío-Luzcando, Joe Carrión-Jumbo. Data Lake Management for Educational Analysis. URL: <https://ieeexplore.ieee.org/document/9935771>

References

1. Zabalawi, I., Kordahji, H., Aftimos, S. Digital Transformation in Digital Transformation in Universities: Strategic Universities: Strategic Framework, Implementation Framework, Implementation Tools, and Leadership. URL: Tools, and Leadership. URL: https://doi.org/10.1007/978-3-031-70779-7_8
2. Michaela Morgan. Achieving Data Governance in Higher Education. URL: Governance in Higher Education. URL: <https://panteon.io/blog/data-governance-in-higher-education>
3. An overview of the role of data warehouses in higher education IT systems. URL: <https://lumendata.com/blog/an-overview-of-the-role-of-data-warehouses-in-higher-education-it-systems/>
4. Diana Martinez-Mosquera, Verónica Beltrán; Diego Riofrío-Luzcando, Joe Carrión-Jumbo. Data Lake Management for Educational Analysis. URL: <https://ieeexplore.ieee.org/document/9935771>