



Лебідь Ярослава Олександрівна,
 магістр спеціальності 122 “Комп’ютерні науки”
 Західноукраїнський національний університет

Ліп’яніна-Гончаренко Христина Володимирівна
 Кандидат технічних наук, доцент.,
 доцент кафедри інформаційно-обчислювальних систем і
 управління,

Західноукраїнський національний університет



**СПИСОК ВИКОРИСТАНИХ
 Джерел**



1. What is a Data Lake? Definition, Architecture, Tools, and Applications | Business and Industry News, Analysis and Expert Insights | Spiceworks. URL: <https://www.spiceworks.com/tech/cloud/articles/what-is-data-lakes/> (дата звернення: 20.11.2022).
2. What is a data lake?. Amazon Web Services, Inc. URL: <https://aws.amazon.com/ru/big-data/datalakes-and-analytics/what-is-a-data-lake/> (дата звернення: 15.11.2022).

Зберігання даних громадян міста

Місто генерує великі обсяги даних від своїх споживачів, операцій і процесів. Науковці даних можуть аналізувати отримані дані, щоб покращити рівень інфраструктури міста. Враховуючи, наскільки «великими» можуть стати дані, озера даних є критично важливими, оскільки вони забезпечують платформу, на якій дані можна зберігати ефективно, безпечно та легко отримувати для аналізу. З розвитком технологій зростає важливість і попит на зберігання цих даних. Ось тут і з'являється архітектура (Рис.1) озера даних.

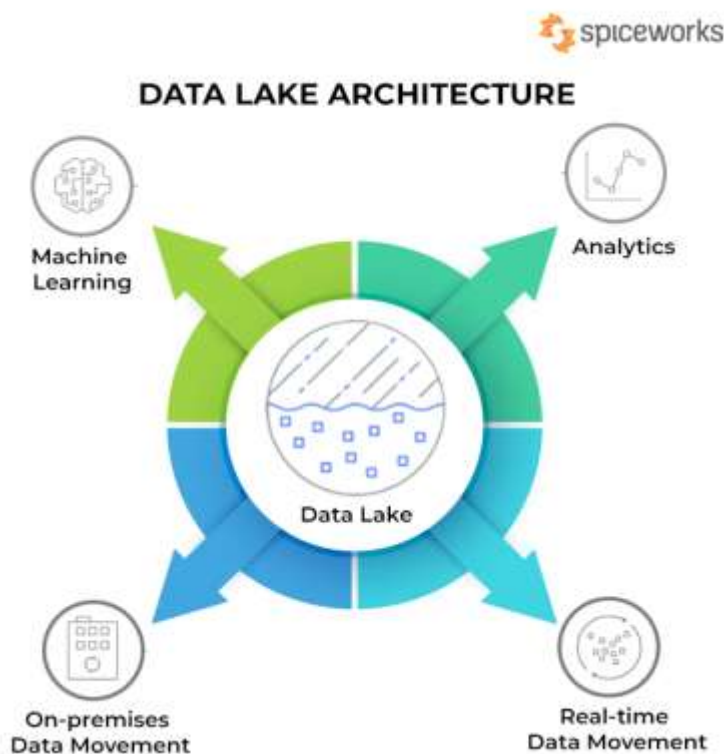


Рис. 1. Архітектура озера даних [1]

Озеро даних визначається як система даних, призначена в основному для неструктурованих даних, де можна зберігати інформацію в одиницях зберігання об'єктів або «блобах» для забезпечення аналітики, машинного навчання та інших видів використання даних в іншому місці ex-situ .

Озера даних забезпечують сховище для зберігання величезних обсягів даних у структурованих, напівструктурованих і неструктурованих формах даних. Вони дозволяють безпечно та ефективно зберігати необроблені дані без фіксованих обмежень для майбутніх аналітичних характеристик.

На відміну від сховищ даних, дані, що зберігаються в озерах даних, не потребують попередньої обробки. Озера даних також забезпечують економічно ефективний спосіб зберігання даних. Дані, що зберігаються в сховищах, використовуються в основному бізнес-аналітиками, тоді як озера даних можуть використовувати науковці, розробники даних і бізнес-аналітики.

Сучасні озера даних забезпечують недорогі сховища даних, які можна масштабувати, зберігаючи свої дані в хмарах, на відміну від традиційних, які використовують локальні сховища. Сучасні озера даних часто мають хмарний аналітичний рівень, який оптимізує продуктивність запитів щодо даних у сховищі даних. Це забезпечує більш ефективну аналітику.

Озера даних надають рішення для організацій, які хочуть накопичувати всі дані з різних джерел даних в одному місці, щоб генерувати з них розуміння. Озера даних дозволяють інструментам бізнес-аналітики безпосередньо отримувати дані, коли це необхідно, що робить їх додатковим інструментом аналітичної підтримки. Вони також підтримують більш швидке надсилання запитів, оскільки дослідники даних можуть виконувати аналітичні запити незалежно від виробничого середовища. Озера даних також добре масштабуються та підтримують багато мов.

Однією з головних проблем, з якими стикаються озера даних, є те, що, оскільки необроблені дані зберігаються «як є», вони повинні мати спеціальні процеси для каталогізації та захисту даних від помилок, до яких запити можуть не виконуватися. Існує також загальний брак навичок, необхідних для використання інструментів в озерах даних, що може вимагати нового найму або внутрішнього професійного розвитку, що призведе до збільшення витрат.

