

УДК 004.67:005

*Ткачук Н. О., студентка 3 курсу  
спеціальності 113 «Прикладна  
математика»*

*Січко Т. В., к.т.н., доцент,  
доцент кафедри інформаційних  
технологій*

## **ЗАСТОСУВАННЯ BIG DATA У БІЗНЕСІ**

*Донецький національний університет імені Василя Стуса, м. Вінниця*

Кожного дня з'являються нові поняття, які кардинально змінюють світ: імітаційне моделювання, штучні нейронні мережі, штучний інтелект, краудсорсинг та старлінк. Цифрові технології постійно розвиваються: створюються нові мови програмування чи вдосконалюються старі, розширюються можливості використання баз даних, з'являються нові функції у текстовому редакторі, тощо. У роботі пропонується розглянути поняття «Big Data».

Концепція «великих даних» не нова, вона виникла в часи мейнфреймів та пов'язаних з ними наукових комп'ютерних обчислень, оскільки наукомісткі обчислення завжди відрізнялись складністю і зазвичай нерозривно пов'язані з необхідністю обробки великих обсягів інформації. Автором терміну «великі дані» є Кліффорд Лінч, редактор журналу «Nature», який зібрав матеріали про явище вибухового зростання обсягу та різноманітності даних та підготував у вересні 2008 року спеціальний випуск журналу, де показав феномен великих даних; термін був запропонований за аналогією з подібними в діловому англійському середовищі метафорами «велика нафта» чи «велика руда» [1].

Великі дані (Big Data) – позначення структурованих і неструктурованих даних величезних обсягів і значного розмаїття, що піддаються ефективній обробці програмних інструментів. З'явилися у кінці 2000-х років із альтернативних традиційних систем управління базами даних і рішення класу Business Intelligence [2].

Визначальними характеристиками для великих даних є ті, які підкреслюють складність задачі обробки і аналізу цих даних. Це так звані 3«V» – volume, velocity, variety, що переводиться, як фізичний об'єм, швидкість приросту даних і необхідність їх швидкої обробки, здатність обробляти дані різних типів. Цей набір даних був розроблений компанією Meta Group у 2001 році з метою вказати на рівну значимість управління даними за всіма трьома аспектами. Згодом з'явилась ще четверта «V» – veracity, тобто достовірність. Через деякий час було створено інтерпретацію з п'яти: viability – життєздатність і value – цінність; семи: variability – змінність та visualization – візуалізація. Але компанія IDC інтерпретує саме четверте V як value (цінність), підкреслюючи економічну доцільність обробки великих об'ємів даних у відповідних умовах.

Розглянемо основні принципи роботи з великими даними.

1. Горизонтальна масштабованість. Це базовий принцип обробки великих даних. Як вже було зазначено, великих даних з кожним днем стає все більше. Відповідно, необхідно збільшувати кількість обчислювальних вузлів, за якими розподіляються ці дані, причому обробка має відбуватись без погіршення продуктивності.

2. Відмовостійкість. Цей принцип витікає з попереднього. Оскільки обчислювальних вузлів у кластері може бути багато (іноді десятки тисяч) та їх кількість, не виключено, буде збільшуватись, зростає ймовірність виходу машин з ладу. Методи роботи з великими даними мають враховувати ймовірність таких ситуацій і передбачати превентивні заходи.

3. Локальність даних. Оскільки дані розподілені по великій кількості обчислювальних вузлів, то, якщо вони фізично знаходяться на одному сервері, а обробляються на іншому, витрати на передачу даних можуть бути невиправдано великими. Тому обробку даних бажано проводити на тій же машині, на якій вони зберігаються [3].

Отже, великі дані – джерело інновацій. Тому бізнес та технічні лідери використовують їх, адже за допомогою великих даних реально не тільки зміцнити свої позиції на ринку, але й запропонувати своїм користувачам клієнтоорієнтований та приємний сервіс.

Розглянемо декілька сфер використання великих даних [4].

1. Фінанси. Гарним прикладом аналітики великих даних у сфері фінансів є компанія «Kreditech», що надає онлайн кредити. Маючи стандартні дані про клієнтів, компанія також використовує інформацію з їхніх постів у соціальних мережах, геолокаційну інформацію та дані про покупки в інтернеті. Потім програма на основі штучного інтелекту обробляє ці дані та визначає, чи існують потенційні ризики надання тому чи іншому клієнтові кредиту – і все це за лічені хвилини.

2. Роздрібна торгівля. «Amazon» – рітейл-гігант, теж широко використовує великі дані. Компанія зберігає понад 1,000,000,000 GB даних на своїх серверах. Ця інформація використовується у багатьох бізнес-процесах, наприклад, для надання покупцям релевантних рекомендацій. «Amazon» відстежує, на які товари покупці дивляться та які врешті купують, і надсилає їм персоналізовані рекомендації щодо майбутніх покупок. Таким чином, близько 35% прибутку компанії складається саме з таких замовлень на основі рекомендацій.

3. Аграрна галузь. «John Deere» – один зі світових лідерів у даній сфері, створив цілу екосистему, яка поєднує обладнання, що оснащене датчиками та хмарним порталом. Ця система відстежує активність у режимі реального часу, аналізує продуктивність та приймає рішення щодо того, що, де та коли саджати.

4. Медіа та розваги. Найкращий приклад використання великих даних у медіа компанії «Netflix». За допомогою машинного навчання, компанія аналізує вподобання своїх глядачів та пропонує їм відповідний контент – 75% відсотків переглядів на Netflix забезпечують персоналізовані рекомендації від

платформи. Таким чином задовольняються потреби понад 195 мільйонів підписників [3].

Отже, великі дані відіграють значну роль у будь-якій індустрії та мають велику низку переваг: від вдосконалення користувацького досвіду до нових потоків прибутку через оцінку ефективності цілих організацій; дають можливість отримати практичну інформацію та дізнатись приховані закономірності.

#### *Список літературних джерел*

1. Великі дані [Електронний ресурс]. Режим доступу: [https://uk.wikipedia.org/wiki/%D0%92%D0%B5%D0%BB%D0%B8%D0%BA%D1%96\\_%D0%B4%D0%B0%D0%BD%D1%96](https://uk.wikipedia.org/wiki/%D0%92%D0%B5%D0%BB%D0%B8%D0%BA%D1%96_%D0%B4%D0%B0%D0%BD%D1%96)
2. Степанюк О.С., Січко Т.В. Особливості використання реляційних та нереляційних баз даних в Big Data. Комп'ютерні технології обробки даних: матеріали всеукр. наук.-практ. конф., м. Вінниця, 2020. С. 103-106.
3. Big Data [Електронний ресурс]. Режим доступу: <https://www.it.ua/knowledge-base/technology-innovation/big-data-bolshie-dannye>
4. Що ж таке Big Data: як бізнес їх використовує [Електронний ресурс]. Режим доступу: <https://www.intellias.ua/blog/big-data-in-business-cases>

**УДК 330.4:004.22**

*Шафорост В. В., Корнієнко К. К.,  
студенти 3 курсу спеціальності 122  
«Комп'ютерні науки»  
Ніколюк П. К., професор, доктор  
фізико-математичних наук.*

## **РОЗРОБКА БАЗ ДАНИХ ТА ЗНАНЬ. BIG DATA**

*Донецький національний університет імені Василя Стуса*

Великі дані «Big Data» – цей термін означає величезний масив інформації, що є структурованою або неструктурованою, а також методи її обробки та аналізу.

Джерела Big Data:

- Мережа Internet – сайти, соціальні мережі, тощо.
- Корпоративна інформація – дані про обсяги закупівель та продажів, транзакції, архіви.

- Показники приладів

Завдання Big Data:

1. Управління великими обсягами інформації, що оновлюється та її подальше зберігання.
2. Обробка актуальної інформації та її структурування.
3. Машинне навчання, аналітика [1].