

УДК 351.004.9

*Підруцький Д.А., Поліщук Д.О.,
здобувачі 2 курсу спеціальності
«Комп'ютерні науки»
Зелінська О.В., доцент
кафедри інформаційних технологій*

АНАЛІЗ ДАНИХ ВЕЛИКИХ ОБ'ЄМІВ (BIG DATA)

Донецький національний університет імені Василя Стуса

На сьогоднішній день існує проблема надмірного накопичення даних у надзвичайно великих розмірах, що зумовлено їх щоденним продукуванням, тож для її вирішення було створено поняття Big Data.

В загальному плані Big Data, вони ж великі дані, представляють з себе позначення структурованих і неструктурованих даних, просто не передати словами, яких величезних обсягів і такої ж значної різноманітності, але ефективно оброблених горизонтально масштабованими програмними інструментами. Якщо підсумувати, то це інформація, що не піддається обробці класичними способами через її величезний об'єм[1].

Big Data згідно правила «П'яти V» діляться на п'ять ключових характеристик [2]:



Рисунок 1. Ключові риси Big Data

Згідно Рис.1, першою рисою є **Volume** – об'єм накопиченої бази даних, який охоплює настільки великий обсяг інформації, що його практично нереально обробляти та зберігати традиційними способами. Для яких потрібен інший підхід та інструменти.

Другою є **Velocity** – швидкість та темпи накопичення і обробки даних постійно збільшуються. Однак останнім часом так само збільшується і попит на

такі технології, які дозволяють використовувати обробку даних в режимі реального часу.

Наступною є різноманітність **Variety**, яка позначає як такі можливості одночасно обробляти неструктуровану та структуровану інформацію. Структурована складає тільки 5 частину від усієї інформації і представляє з себе – дані, які можна класифікувати, наприклад інформацію з банківської бази даних, де чітко вказаний перелік клієнтів та їхні фінансові транзакції. Неструктурована представляє всі інші 80% даних і охоплює різноманітні масиви даних, такі як фото, відео і інші, наприклад з соціальних мереж.

Також потрібно перевіряти дані на справжність, тому далі йде **Veracity** – виокремлення достовірних даних. Якість отриманої інформації може відрізнятися, тим самим впливаючи на точний аналіз.

І протилежність минулому **Variability** – невідповідність інформації, яка ускладнює та іноді сильно заважає процесам обробки та управління інформації.

З самим представленням поняття Big Data та його характеристиками ми проаналізували, а зараз представимо сфери їх застосування.

Відповідно статті про бази даних [3], на разі Big Data доволі широко застосовують в галузі ритейлу (продажі), де вони допомагають різноманітним компаніям знаходити вигідніші території для цільових клієнтів і визначати, де існує необхідність у нових філіалах цих компаній. До прикладу, такий підхід використовує мережа магазинів «АТБ».

Крім всякого роду компаній, Big Data знаходить своє місце також в багатьох сферах, серед яких є такі як вище згаданий ритейл, медицина, банківська справа і фінанси, промисловість, енергетика, туризм, екологія, розваги. А саме завдяки обробці та аналізу великого масиву даних, представники бізнесу, влади, науки, розробники програмних забезпечень та інші зацікавлені особи покращують якість товарів і послуг, розвиваючи при цьому свій бізнес.

У наслідок розвитку технологій у світі значно зросла кількість даних. Технології Big Data дають змогу швидко і якісно отримувати користь із цього потоку інформації. З їх допомогою державні структури і представники бізнесу оптимізують різні процеси, а кінцеві споживачі отримують більш якісні послуги.

Список літератури

1. WIKIPEDIA URL: (дата звернення: 02.12.2022)
https://en.wikipedia.org/wiki/Big_data
2. GAGROUP URL: (дата звернення: 02.12.2022)
<https://qagroup.com.ua/publications/shcho-take-big-data-i-iaak-tce-pratciuiie/>
3. AICONFERENCE URL: (дата звернення: 02.12.2022)
<https://aiconference.com.ua/uk/news/tehnologii-big-data-klyuchevie-harakteristiki-osobennosti-i-preimushchestva-97883>