

BIG DATA ТА АНАЛІТИКА

Вінницький Національний Технічний Університет

Анотація

У доповіді розглядаються основні аспекти Big Data та аналітики. Обговорюються характеристики Великих Даних, такі як обсяг, швидкість, різноманітність, достовірність та цінність. Представлено сучасні технології для обробки та зберігання даних бази даних та об'єктні сховища. Також розглядаються типи аналітики даних.

Ключові слова: Big data, аналітика, бази даних, безпека даних, інновації.

Abstracts.

The report covers the main aspects of Big Data and analytics. It discusses the characteristics of Big Data, such as volume, speed, variety, reliability and value. Modern technologies for processing and storing data, such as databases and object warehouses, are presented. The types of data analytics are also discussed.

Keywords: Big data, analytics, databases, data security, innovations.

Вступ

У сучасному світі обсяг даних зростає з неймовірною швидкістю завдяки розвитку цифрових технологій, соціальних медіа, мобільних пристроїв та інтернету речей. Великі Дані стали невід'ємною частиною бізнесу, науки, медицини та багатьох інших галузей. Вони відкривають нові можливості для аналізу та прийняття обґрунтованих рішень. Аналітика даних дозволяє організаціям отримувати цінні інсайти, підвищувати ефективність, оптимізувати процеси та створювати інноваційні продукти та послуги.

Визначення та технології Big data

Big Data (укр. великі дані) — це великий масив структурованої та неструктурованої інформації, а також інструменти, підходи, методи обробки та зберігання даних. Важливість великих даних залежить не тільки від їхньої кількості, а й від того, як компанія їх інтерпретує та використовує. Через об'єм та різноманітність даних обробляти їх традиційним програмним забезпеченням неможливо. Наприклад, оцінити попит на самокати та велосипеди в одному магазині впродовж року людина може самотійно. Проте якщо потрібно проаналізувати такий попит у сотнях магазинів кількох країн, власнику торговельної мережі потрібна буде допомога новітніх Big Data-інструментів.[1]

Big Data характеризується трьома основними атрибутами, відомими як "3V":

–Обсяг (Volume): Кількість даних, що генеруються, зберігаються та аналізуються. Вимірюється в петабайтах та екзабайтах.

–Швидкість (Velocity): Швидкість, з якою дані генеруються та обробляються. Це може бути потік даних у реальному часі.

–Різноманітність (Variety): Різні типи даних, які можуть бути структурованими, неструктурованими та напівструктурованими (тексти, зображення, відео, сенсорні дані тощо).

Обробка даних

Для роботи з великими обсягами даних використовуються різноманітні технології та платформи. Деякі з них включають:

- Hadoop: Відкрита програмна платформа для зберігання та обробки великих обсягів даних на розподілених кластерах.
- Spark: Платформа для обробки даних в пам'яті, яка забезпечує високу швидкість обробки.

Зберігання даних

NoSQL бази даних: Наприклад, MongoDB, Cassandra, які розроблені для зберігання великих обсягів неструктурованих даних.

Об'єктні сховища: Amazon S3, Google Cloud Storage, які забезпечують масштабоване зберігання даних.

Аналітика даних

Аналітика даних - процес збирання та обробки інформації з метою отримання правильних висновків на основі аналітичних та логічних міркувань. Основна мета аналітики даних з нуля – пошук сенсу та способів застосування отриманих знань на практиці.[2]

Аналітика даних включає різні методи та інструменти для аналізу великих обсягів даних. Вона поділяється на декілька типів:

- Описова аналітика (Descriptive Analytics): Аналіз історичних даних для виявлення тенденцій та шаблонів.
- Діагностична аналітика (Diagnostic Analytics): Аналіз даних для розуміння причин подій.
- Прогнозна аналітика (Predictive Analytics): Використання статистичних моделей та алгоритмів машинного навчання для прогнозування майбутніх подій.
- Прескриптивна аналітика (Prescriptive Analytics): Рекомендації щодо дій на основі прогнозів.

Інструменти аналітики

- R та Python: Мови програмування з потужними бібліотеками для аналізу даних (pandas, NumPy, scikit-learn).
- Tableau, Power BI: Інструменти для візуалізації даних та створення інтерактивних звітів.
- Machine Learning платформи: TensorFlow, PyTorch для створення моделей машинного навчання.

Виклики та можливості

Виклики

- Безпека та конфіденційність: Захист даних від несанкціонованого доступу.
- Якість даних: Забезпечення точності та повноти даних.
- Сумісність: Інтеграція різних джерел даних та систем.

Можливості

- Прийняття рішень на основі даних: Підвищення ефективності бізнес-процесів.
- Інновації: Розробка нових продуктів та послуг на основі аналізу даних.
- Персоналізація: Надання індивідуальних пропозицій клієнтам на основі їх поведінки та вподобань.

Висновок

Big Data та аналітика є потужними інструментами для сучасних організацій, що дозволяють їм отримувати цінні інсайти та приймати обґрунтовані рішення. Незважаючи на численні виклики, пов'язані з обробкою та зберіганням великих обсягів даних, потенціал для розвитку бізнесу та інновацій є величезним. Впровадження передових технологій та методів аналізу дозволяє організаціям залишатися конкурентоспроможними в сучасному світі.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Що таке Big Data? [Електронний ресурс] – Режим доступу: <https://hub.kyivstar.ua/articles/shho-take-big-data>
2. Аналітика Даних: Як і Навіщо? [Електронний ресурс] – Режим доступу: <https://sparsim.org/analitika-danykh-yak-i-navishcho/>

Березняк Максим Володимирович – студент групи 2БС-22б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький Національний Технічний Університет, Вінниця, e-mail: maksimbereziniak21@gmail.com

Berezniak Maksym Volodymyrovych - student of group 2BS-22b, Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: maksimbereziniak21@gmail.com