

Аналіз та Організація Big Data

Руслан Селезько

Кафедра інфокомунікаційної інженерії ім. В.В. Поповського
Харківський національний університет радіоелектроніки
Харків, Україна
ruslan.selezko@nure.ua

Big Data Analysis and Organization

Ruslan Selezko

dept. of Infocommunication Engineering V.V. Popovsky
Kharkiv National University of Radio Electronics
Kharhiv, Ukraine
ruslan.selezko@nure.ua

Анотація—У статті розглянуто значення та потенціал великих даних (BD) як важливого технологічного прориву в академічних і бізнес-спільнотах. Великі дані описуються як обсяги різноманітних даних, що генеруються, захоплюються і оброблюються з великою швидкістю, важко піддаються обробці існуючими технологіями. Виокремлено потенціал великих даних для розвитку інноваційних інсайтів, продуктів та послуг у різних сферах, таких як електронна комерція, наука і охорона здоров'я. Значна увага приділяється також аналітиці великих даних (BDA), яка стає популярною практикою для побудови цінної інформації з BD та використання її для покращення операційної ефективності та досягнення конкурентних переваг. Автор аналізує існуючі дослідження з метою ідентифікації різних викликів BD та методів BDA, щоб підтримати розвиток академічної спільноти.

Abstract—The article examines the importance and potential of Big Data (BD) as an important technological breakthrough in the academic and business communities. Big data is described as volumes of various data that are generated, captured and processed at a high speed, difficult to process with existing technologies. The potential of big data for the development of innovative insights, products and services in various fields, such as e-commerce, science and health care, is highlighted. Considerable attention is also paid to big data analytics (BDA), which is becoming a popular practice to build valuable information from BD and use it to improve operational efficiency and achieve competitive advantage. The author analyzes existing research to identify the various challenges of BD and BDA methods to support the development of the academic community.

Ключові слова – великі дані; інновації; виклики; методи аналізу даних; розвиток; потенціал; безпека даних.

Keywords – big data; innovations; challenges; data analysis methods; development; potential; data security.

I. ВСТУП

У сучасному цифровому світі, де кожна діяльність і кожен аспект життя стають все більш цифровими, великі дані (BD) відіграють ключову роль у вирішенні проблем, що стоять перед науковими та бізнес-спільнотами. BD описуються як обсяги різноманітних даних, що

генеруються, захоплюються та оброблюються з великою швидкістю. Ці дані важко піддаються обробці існуючими технологіями, і вимагають нових методів та інструментів для аналізу та використання.

У 2008 році термін «великі дані» вперше з'явився в пресі, коли редактор журналу Nature, Кліффорд Лінч, опублікував статтю про майбутній розвиток науки за допомогою технологій обробки великих обсягів даних. До 2009 року цей термін використовувався переважно в наукових дослідженнях, але після видання кількох інших статей преса почала активно вживати термін «Big Data», що вживається й досі.

У 2010 році почали з'являтися перші спроби вирішення проблеми великих обсягів даних, з'явилися програмні продукти, спрямовані на зменшення ризиків при роботі з великими обсягами інформації. До 2011 року великі компанії, такі як Microsoft, Oracle, EMC і IBM, проявили зацікавленість у використанні великих даних у своїх стратегіях розвитку і досягли значних успіхів. У 2013 році університети почали вивчати великі дані як окремий предмет, залучаючи до цього не лише спеціалістів з обробки даних, а й інженерів разом із комп'ютерними науками.

Великі дані включають в себе не тільки структуровані дані, такі як дані з баз даних, але й неструктуровані дані, такі як текстові файли, зображення, аудіо та відео. За допомогою сенсорів, сучасні пристрої генерують величезні обсяги даних з різних джерел, включаючи соціальні мережі, мобільні додатки, сенсори IoT та інші. [1-2]

Одним з ключових викликів, що виникає з великих обсягів даних, є їхня обробка. Традиційні методи обробки даних не завжди ефективні в умовах великих обсягів даних. Тому, розвиток нових методів та технологій для обробки даних стає важливим завданням. Наприклад, технології розподіленого обчислення, які базуються на кластерах серверів, дозволяють обробляти великі обсяги даних паралельно і ефективно.

II. ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Дисципліна дослідження великих даних (BD) та аналітики великих даних (BDA) все ще розвивається і ще не встановлена, тому повне розуміння явища, його визначення та класифікація ще не були повністю визначені. Існуючий прогрес у галузі BD та BDA показав не лише відсутність досліджень управління у цій сфері, але й відсутність теоретичних конструкцій та академічної строгості – можливо, як результат методологічного, а не академічного виклику. У цілому також відчутний брак досліджень, які комплексно вирішують ключові виклики BD або досліджують можливості для нових теорій чи емерджентних практик. Таким чином, існує потреба у підсумовуванні викликів BD та пов'язаних методів BDA для можливості вказівників. [1-3]

Big Data мають 3 ключові характеристики, які також відомі як "3V":

1. **Об'єм (Volume):** описує надзвичайно великий обсяг даних, які є складними для традиційної обробки та зберігання. Для їх ефективного управління потрібен новий підхід та вдосконалені інструменти.
2. **Швидкість (Velocity):** характеристика вказує на швидкість накопичення даних, що постійно зростає. Це також означає потребу у швидкій обробці даних, в тому числі в режимі реального часу.
3. **Різноманітність (Variety):** можливість обробки як структурованої, так і неструктурованої інформації. Наприклад, це може бути інформація з баз даних та дані з соціальних мереж.

До основних трьох, також можна додати:

1. **Достовірність (Veracity):** оскільки обсяг даних постійно зростає, важливо визначити достовірність даних, оскільки якість даних може суттєво впливати на результати аналізу.
2. **Мінливість (Variability):** невідповідність та мінливість даних ускладнюють їх обробку та управління.

Ці характеристики підкреслюють складність та масштабність завдань, пов'язаних з аналізом великих обсягів даних, а також необхідність вдосконалення методів та інструментів для ефективного використання цих даних. [2-10]

Використання великих даних має потенціал для розвитку інноваційних інсайтів, продуктів та послуг у різних сферах. У науці, аналіз великих даних дозволяє дослідникам виявляти нові закономірності та знання у великих обсягах даних. Наприклад, у геноміці, аналіз великих обсягів геномних даних допомагає виявляти генетичні маркери для різних захворювань.

У бізнесі, великі дані допомагають підприємствам розуміти поведінку споживачів, прогнозувати попит на товари та послуги, оптимізувати ланцюжки постачання та покращувати стратегії маркетингу. Наприклад, аналіз

великих даних у роздрібній торгівлі дозволяє роздрібним мережам прогнозувати попит на товари та планувати запаси, що дозволяє їм зменшити втрати через непродані товари та збільшити прибуток. [3]

Аналітика великих даних (BDA) – це процес виявлення, інтерпретації та використання великих обсягів даних з метою отримання цінної інформації, яка може бути використана для прийняття стратегічних рішень та виявлення нових можливостей. BDA використовує різноманітні методи, алгоритми та інструменти, такі як статистичний аналіз, машинне навчання, штучний інтелект та інші, для аналізу великих обсягів даних та виявлення в них закономірностей та тенденцій.

Незважаючи на потенціал, BDA стикається з викликами, такими як проблеми з конфіденційністю та безпекою даних, неоднорідність даних та складність аналізу великих обсягів даних. Однак, розвиток нових технологій та методів, таких як блокчейн, розподілені системи та квантові обчислення, може допомогти вирішити ці виклики та розширити можливості використання BDA у майбутньому.

Хоча технології обчислень, необхідні для обробки таких даних, розвиваються, людська експертиза та можливості, необхідні для ефективного використання BD, відстають. Широкі виклики BD можна розділити на три основні категорії згідно з життєвим циклом даних: виклики даних, процесу та управління. [1-3]

Кроки аналізу великих даних:

- Збір Даних (перший крок у процесі BDA – це збір великих обсягів даних з різних джерел, включаючи бази даних, веб-сайти, сенсори IoT, соціальні мережі та інші джерела.)
- Підготовка Даних (дані, отримані з різних джерел, можуть бути неоднорідними та неповними. Тому перед аналізом вони потребують підготовки, що включає очищення, перетворення та інтеграцію даних.)
- Аналіз Даних (після підготовки даних проводиться їх аналіз за допомогою різних методів та інструментів. Це може включати в себе статистичний аналіз, візуалізацію даних, машинне навчання та інші техніки.)
- Використання Результатів (останній крок – використання отриманих результатів для прийняття рішень, виявлення нових можливостей та покращення бізнес-процесів.)

Перший крок у процесі аналітики великих даних полягає у зборі різноманітних даних з різних джерел, таких як бази даних, веб-сайти, датчики IoT, соціальні мережі та інші. Після збору даних вони потребують підготовки, включаючи очищення від шуму та аномалій, інтеграцію різних джерел даних і перетворення їх у формат, придатний для аналізу. [3-8]

На етапі аналізу даних використовуються різноманітні методи та інструменти для виявлення зв'язків,

закономірностей та тенденцій у даних. Це включає в себе статистичний аналіз, машинне навчання, штучний інтелект та інші техніки. Наприклад, за допомогою машинного навчання можна побудувати прогностичні моделі для прогнозування майбутніх подій на основі історичних даних. [3-5]

Важливим етапом аналізу великих даних є їх візуалізація. Вона дозволяє представити отримані результати у вигляді графіків, діаграм, карт та інших візуальних засобів. Візуалізація даних допомагає зрозуміти складні зв'язки та виявити важливі тенденції, що можуть бути важливі для прийняття стратегічних рішень.

Найважливішим же етапом аналізу великих даних є виявлення інсайтів – нових, неочевидних знань або відкриття, що можуть бути корисними для прийняття рішень або розробки нових продуктів та послуг. Інсайти можуть бути виявлені шляхом виявлення невідомих закономірностей, прогнозування майбутніх тенденцій або виявлення нових можливостей для оптимізації процесів.

Отримані інсайти використовуються для прийняття стратегічних рішень у різних сферах діяльності. Наприклад, в бізнесі інсайти можуть використовуватися для розробки стратегій маркетингу, управління ризиками, оптимізації процесів та прийняття рішень про інвестування. У науці вони можуть сприяти виявленню нових наукових відкриттів та розробці нових технологій.

Процес аналізу великих даних є постійною ітеративною діяльністю, яка вимагає постійного вдосконалення методів та інструментів, постійне вдосконалення дозволяє підтримувати конкурентоспроможність та використовувати нові можливості, які виникають у зв'язку зі зміною умов діяльності. [7-10]

III. ВИСНОВКИ

У наш час концепція Big Data є важливим соціально-економічним явищем, що виникає внаслідок появи нових технологічних можливостей для аналізу величезних обсягів даних. Швидкий розвиток інформаційних технологій протягом останніх десятиліть, що виник внаслідок інформаційної революції, змінив підходи до збору та управління маркетинговими даними.

Великий обсяг різноманітних та неструктурованих даних, що надходять у цифровому середовищі та мають велике значення для компаній, потребує інтелектуального

підходу до їх збору, обробки та аналізу, що створює потребу у автоматизації цих процесів. Поєднання інструментів, спрямованих на задоволення потреб сучасних організацій у перетворенні хаотичних даних на корисну інформацію для прийняття рішень, утворює поняття Big Data. Технологія Big Data має велике значення в маркетинговій сфері, оскільки дозволяє компаніям отримувати детальну інформацію про своїх клієнтів окремо, що сприяє переходу до персоналізованого обслуговування та автоматичного формування індивідуальних маркетингових програм. Це полегшує підвищення ефективності маркетингових інвестицій шляхом надання споживачам більш релевантної інформації. Для зарубіжних компаній використання Big Data вже стало стандартом, необхідним для забезпечення конкурентоспроможності.

Перехід українських компаній до управління на основі інтелектуального аналізу даних потребує часу та вирішення таких проблем, як висока вартість програмного забезпечення, адаптація спеціалістів до нових технологій та технічні перешкоди, пов'язані з нескоординованістю використовуваних алгоритмів.

ЛІТЕРАТУРА REFERENCES

- [1] M. Kowalczyk. Big Data and Information Processing in Organizational Decision Processes. Research Paper Published: 08 August 2014. Volume 6, pages 267–278, (2014)
- [2] S. Fosso Wamba, S. Akter, A. Edwards. How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. International Journal of Production Economics. Volume 165, July 2015, Pages 234-246
- [3] R.J. Kauffman et al. Business and data analytics: new innovations for the management of e-commerce. Electron. Commer. Res. Appl. (2012)
- [4] Columbus L. Ten Ways Big Data Is Revolutionizing Marketing And Sales [Електронний ресурс] / Louis Columbus // Forbes. – 2016.
- [5] Кравчук С. Що таке Big Data [Електронний ресурс] / Сергій Кравчук // The Future. – 2017.
- [6] Goldfein J. Big Data & Marketing: Value, Problems, and Solutions [Електронний ресурс] / Joshua Goldfein. – 2016.
- [7] Wang, Y., & Wiebe, V. J. (2014). Big Data Analytics on the characteristic equilibrium of collective opinions in social networks. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 8(3), 29–44.
- [8] Yi, X., Liu, F., Liu, J., & Jin, H. (2014). Building a network highway for big data: architecture and challenges. IEEE Network, 28(4), 5–13.
- [9] Zhang, F., Liu, M., Gui, F., Shen, W., Shami, A., & Ma, Y. (2015a). A distributed frequent itemset mining algorithm using Spark for Big Data analytics. Cluster Computing, 18(4), 1493–150