

Великі дані як інформаційна технологія

Наталя Шаховська, Володимир Москалюк

Кафедра інформаційних систем та мереж, Національний університет “Львівська політехніка”, УКРАЇНА, м. Львів, вул. С. Бандери, 12,
E-mail: natalya233@gmail.com

Abstract – This paper introduced the concept of the term Big Data and analyzes the cause of their appearance.. We describe the data space as the technology of working with Big data.

Ключові слова – Великі дані, простір даних, методи аналізу даних, паралельне опрацювання даних.

I. Вступ

Опрацювання інформаційних ресурсів, що використовують різні моделі даних, схеми керування тощо вимагає розроблення уніфікованого методу доступу до них для того, щоб надати можливість користувачу вибирати адекватний інструментарій для вивчення та використання різних засобів опрацювання даних. Необхідність у цьому виникає в організацій, робота яких полягає в опрацюванні великої кількості різнотипних, взаємозалежних джерел даних, для яких не всі семантичні взаємозв'язки відомі і вказані. У деяких випадках семантичні зв'язки невідомі через невизначену кількість початкових джерел або через брак кваліфікованих людей у визначенні таких зв'язків. У інших випадках, не всі семантичні зв'язки необхідні для класифікації послуг користувачам. Тому в користувачів немає єдиної схеми, за якою вони можуть створювати запити відносно цільових задач.

За усієї важливості відомих результатів, теоретичні та експериментальні дослідження повинні розвиватися в напрямку: розроблення ефективних засобів опрацювання даних з різнотипних інформаційних ресурсів та вироблення засад і критеріїв оцінювання якості інтегрованих даних, які би підвищувати ефективність прийнятих рішень.

II. Основний матеріал

Основними проблемами, які виникають при обробці даних, є відсутність методів аналізу, придатних до застосування через їх різнотипність, потреба у значних людських ресурсах для підтримки процесу аналізу даних, висока обчислювальна складність наявних алгоритмів аналізу та стрімке зростання обсягу зібраних даних. Вони призводять до постійного зростання часу аналізу навіть при регулярному оновленні апаратних засобів серверів, а також – необхідність роботи із розподіленими базами даних, можливості яких

більшість існуючих методів аналізу даних не використовують ефективно. Таким чином, виникає задача розроблення ефективного методу аналізу даних, що може застосовуватись до розподілених баз даних різних предметних областей. Тому доцільно розробляти методи та засоби консолідації даних та використання їх для аналізу.

Великі дані (Big Data) в інформаційних технологіях – набір методів та засобів опрацювання структурованих і неструктурованих різнотипних динамічних даних великих обсягів з метою їх аналізу та використання для підтримки прийняття рішень. Є альтернативою традиційним системам управління базами даних і рішеннями класу Business Intelligence. До цього класу відносять засоби паралельного опрацювання даних (NoSQL, алгоритми MapReduce, Hadoop) [1, 2].

Визначальними характеристиками для Великих даних є обсяг (volume, в сенсі величини фізичного обсягу), швидкість (velocity в сенсах як швидкості приросту, так і необхідності високошвидкісної обробки та отримання результатів), різноманіття (variety, в сенсі можливості одночасної обробки різних типів структурованих і напівструктурованих даних).

Вважається, що Великі дані є технологією Hadoop. Проте Великі дані є дещо більше, ніж Hadoop. Одною з ключових вимог є розуміння і навігація по федеративних джерелах великих даних - щоб виявити дані на місці. Нова технологія підтримує також індекси, пошук та навігацію різних джерел Великих даних.

Простір даних – це блоковий вектор, що містить множину інформаційних продуктів предметної області, поділену на три блоки: структуровані дані (бази, сховища даних), напівструктуровані дані (XML, електронні таблиці) та неструктуровані дані (текст). Над цим вектором та його окремими елементами визначено операції та предикати, які забезпечують: перетворення різних елементів вектора один в одного; об'єднання елементів одного типу; пошук в елементах за ключовим словом.

Висновок

Показано, що для формального подання Великих даних доцільно використовувати простір даних.

Література

1. White, Tom (10 May 2012). Hadoop: The Definitive Guide. O'Reilly Media. p. 3. ISBN 978-1-4493-3877-0.
2. Big Data for Development: From Information- to Knowledge Societies", Martin Hilbert (2013), SSRN Scholarly Paper No. ID 2205145). Rochester, NY: Social Science Research Network; <http://papers.ssrn.com/abstract=2205145>