

УДК 004.9

А.К. Карнаухов, О.О. Кузьо

Тернопільський національний технічний університет імені Івана Пулюя, Україна

## КОНСОЛІДАЦІЯ ІНФОРМАЦІЇ КОРИСТУВАЧІВ ЗА ДОПОМОГОЮ СТРУКТУР BIG DATA

Karnaukhov A.K., Kuzo O.O.

### CONSOLIDATION OF USER INFORMATION USING BIG DATA STRUCTURES

В сучасних умовах відкриваються можливості для вирішення дуже актуальних завдань, пов'язаних з обробкою великих об'ємів даних від користувачів. Особливо перспективною є інтеграція сучасних інформаційних технологій, пов'язаних із збором та обробкою великих неструктурованих даних (Big Data), інтелектуальним аналізом даних (Data Mining), розподіленими базами даних.

Великі дані (Big Data) – це сучасний вектор у сфері інформаційних технологій і систем, що включає перелік технологій, методів, та інструментів для збору та аналізу великих обсягів різноманітних структурованих і неструктурованих даних з метою отримання ефективних результатів в умовах поширення інформації в мережі багатьох обчислювальних вузлів і їх безперервне оновлення.

Аналітику Big Data можна охарактеризувати наступними параметрами: - обсяг, кількість згенерованих даних. Цей показник залежить від того, чи можна вважати масив даних вважати Big Data. Дані традиційно зберігаються на серверах SQL у хмарних середовищах; - різноманітність, категорія, до якої потрапляють Big Data. Розуміння цієї приналежності дозволяє аналітику найбільш ефективно обробляти інформацію; - швидкість, з якою дані генеруються або обробляються для досягнення поставлених цілей; - мінливість, тобто нестабільність на часі даних; - достовірність, тобто якість зібраних даних, від якої залежить точність аналізу; - складність, яка полягає у трудомісткості процесу зв'язування та встановлення зв'язків між даними.

Обробка Big Data в даний час зазвичай передбачає впровадження спеціальних програмних комплексів, таких як Hadoop, які дозволяють обробляти великі обсяги даних на основі концепції Map-Reduce. Hadoop — це структура, на якій розробляються програми для аналізу та візуалізації великих даних. Зберігання даних у цій структурі здійснюється за допомогою спеціальної розподіленої файлової системи HDFS (англ. Hadoop Distributed File System), яка є основою Hadoop і дозволяє зберігати та отримувати доступ до даних одночасно на кількох вузлах кластера.

Алгоритм Map-Reduce використовується для обробки великих масивів даних, і кожен етап Map має завершити свою відповідну роботу до початку Reduce, а вхідні дані потрібно попередньо обробити. Одним із найактуальніших завдань сучасних інформаційних технологій — швидка обробка великих об'ємів даних. Ефективне рішення кожного завдання має можливість пришвидшити приймання рішень на основі попередньо отриманих даних. Зі збільшенням можливостей збору інформації за допомогою багатьох різних пристроїв розширюється і кількість даних.

### Література

1. Кеннет Кукієр. Великі дані. Як вони змінюють наші уявлення про світ [Електронний ресурс] / Кеннет Кукієр, Віктор Майєр-Шенбергер.
2. James Manyika. Big data techniques and technologies / J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Hung Byers // Big data: The next frontier for innovation, competition, and productivity. - McKinsey Global Institute. -
3. Neil Biehn. The Missing V's in Big Data: Viability and Value.