

3. [https://en.wikipedia.org/wiki/Capital_\(economics\)](https://en.wikipedia.org/wiki/Capital_(economics))
4. <https://www.forumforthefuture.org/project/five-capitals/overview>
5. Kean Birch 1, Les Levidow and Theo Papaioannou. Sustainable Capital? The Neoliberalization of Nature and Knowledge in the European “Knowledge-based Bio-economy” – Sustainability 2010, 2, 13 September 2010. – P. 2899-2918.
6. Neva R. Goodwin. Five Kinds of Capital: Useful Concepts for Sustainable Development. – Global Development and Environment Institute, Working Paper No. 03-07, September 2003. – P. 1-12.

УДК: 330.4:004.22

І.М. Онищенко

Удосконалення методів обробки та зберігання даних за допомогою інструментів "Big Data" та Map Reduce

В умовах нового витку технологічного розвитку та переходу до парадигми інформаційного суспільства, де більшість зайняті виробництвом, зберіганням, обробкою та реалізацією даних, особливого значення набуває концепція “Big Data”. У статті проведено дослідження феномену "Big Data", а також розглянуто питання про те, що слід розуміти під цим поняттям, що є основним джерелом даних та які основні властивості та інструменти формують комплексне поняття "Big Data".

Ключові слова: інформаційні технології в економіці, Big Data, Map Reduce.

В умовах нового витку технологічного розвитку та переходу до парадигми інформаційного

общества, где большинство заняты производством, хранением, обработкой и реализацией данных, особенно значения получает концепция "Big Data". В статье проведено исследование феномена "Big Data", его инструментов и методов, также рассмотрен вопрос о том, что следует понимать под этим термином, что является источником данных, какие основные свойства и инструменты формируют комплексное понятие "Big Data".

***Ключевые слова.** Информационные технологии в экономике, Big Data, Map Reduce.*

The article deals with the concept of "Big Data" as information technology for economic applications. The questions of what is "Big Data", what are the sources, tools and features of "Big Data" are investigated in the article. Next the author describes different approaches to definition the concept of "Big Data" based on the amount of data and modern computing capacity and based on the "Three V" approach. Also application of "Big Data" technologies for data storage, processing and modeling. After review the concepts and features of "Big Data" the author shows an example how "Big Data" with distributed file systems and Map Reduce can improve traditional approach on data processing.

***Keywords.** Information Technologies in Economics, Big Data, Map Reduce.*

Актуальність. Сучасне суспільство переживає черговий бум інформаційних технологій, який цього разу пов'язаний із швидким, експоненціальним зростанням обсягів інформації. При чому частина структурованої інформації зростає не так стрімко. Основна частина

приросту обсягів інформації – це неструктуровані або слабо структуровані дані. Класичні ж методи обробки та зберігання даних не можуть впоратися із такими обсягами. Іншим недоліком є також те, що класичні алгоритми обробки даних орієнтовані на роботу зі структурованими даними та виявляються неефективними при їх застосуванні на задачах з неструктурованими даними.

Для вирішення вказаних вище задач одночасно в кількох найбільших світових компаніях індустрії інформаційних технологій почали розробляти абсолютно нові підходи до проблеми зберігання та обробки інформації з метою отримання корисних знань. Пізніше зусилля окремих компаній були об'єднані у єдиний проект в результаті якого було отримано систему нових інструментів, методів, апаратного та програмного забезпечення для аналізу даних великих обсягів та поганої структурованості. Система таких методів та підходів отримала назву технологій "Big Data".

Треба відзначити, що питання визначення основних властивостей поняття "Big Data" є доволі гострим та актуальним серед спеціалістів та науковців у сфері інформаційних технологій всього світу.

Аналіз останніх досліджень і публікацій. Термін Big Data був введений у 2008 році Кліффордом Лінчем [1], доктором з інформатики університету Берклі. Також слід відзначити роботи В. Майєр-Шенбергера та К. Кук'єра [2], Ж.-П. Дейкса [3], які також проводили фундаментальні дослідження у сфері великих даних.

Хоча сам термін був введений в академічному середовищі, широке застосування та поширення він здобув також у практичних дослідженнях в рамках технологічних проектів передових компаній, зокрема праці Доуг Хеншена

(Oracle), Клінта Фінлі (Microsoft), Агама Шаха (Hewlett-Packard) та інших [4 – 6].

Останнім часом спостерігається підвищення інтересу до даної теми з боку дослідників з України та країн СНД. Так Н. Шаховська та Ю. Болюбаш у своїх працях досліджують сучасні бази даних, які використовують технології "Big Data" [7], Л. Черняк досліджує концепцію "Big Data" в цілому [8], Р. Ускенбаєва досліджують питання впровадження "Big Data" в електронному уряді [9, 10].

Невирішені проблеми. У той же час практичні питання щодо вдосконалення інструментів аналізу, зберігання та обробки даних за допомогою методів "Big Data" та розподілених файлових систем в науковій літературі залишаються недослідженими. Також немає остаточної відповіді на питання що таке "Big Data" і як розуміти цей термін.

Мета статті. У даній статті основною метою є аналіз підходів до визначення концепції "Big Data", джерел формування даних, проблем та рішень які найбільш відповідають сучасному стану та вимогам індустрії. Також будуть досліджені фундаментальні рішення на яких будуються інструменти "Big Data" – модель розподілених обчислень та системи алгоритмів Map Reduce.

Постановка завдання. У даній статті поставлено та розв'язано такі наукові завдання: аналіз підходів до визначення терміну "Big Data", його ознак та властивостей, а також досліджено питання застосування інструментів "Big Data" та Map Reduce та їх переваги у порівнянні із традиційними підходами.

Виклад основного матеріалу.

Термін "Big Data" вперше згадується 2008 року у випуску британського наукового видання Nature, який присвячений темі "Як впливають на майбутнє науки

технології, які відкривають можливості роботи з великими обсягами даних?"[1]. При чому автор статті Кліфорд Лінч пропонує термін "Big Data", який відповідає не стільки кількісному вимірюванню, але переходу від кількості у якість. Термін був запропонований по аналогії до таких метафор як "Велика Нафта" або "Велика руда".

ІТ індустрія швидко почала використовувати термін "Big Data" підкреслюючи роль даних як джерела прихованої інформації. В індустрії вже певний час обговорювалась необхідність розробки технологій для роботи з великими обсягами даних у зв'язку із розповсюдженням сервісів радіочастотної ідентифікації, соціальних мереж тощо.

Визначення 1: "Big Data" – термін, який застосовується до даних настільки великої розмірності або складності, що традиційні методи обробки даних для них не працюють.

Серед проблем, які виникають із "Big Data" виділяють – аналіз, пошук, зберігання, передачу, візуалізацію даних та інші [11].

Суть проблеми "Big Data" заключається в тому, що традиційні методи обробки даних дають недостатньо якісні результати, або не дають їх взагалі. Причини таких збоїв можуть бути різними. Метою розв'язання проблем пов'язаних з обробкою великих обсягів даних є отримання можливості для більш точного прогнозування, систем прийняття рішень, підвищення ефективності операційної діяльності, зниження ризиків та втрат. Досягнення поставленої мети пропонується через побудову ефективних, в умовах "Big Data", методів прогнозування та аналізу, а також інших способів отримання прибутку з даних.

Очевидний недолік даного визначення – недостатньо чітко окреслені критерії належності до означуваного поняття що ж вважати "Big Data", а що ні. Традиційні методи обробки даних також розвиваються і можуть зберігати та обробляти все більші й більші обсяги інформації. Таким чином, задачі які раніше були віднесені до "Big Data" сьогодні до них вже не відносяться відповідно до визначення.

Методологічно більш коректно базувати визначення поняття "Big Data" саме на методах зберігання та обробки даних, які принципово відрізняються від традиційних. Відповідно, приходимо до наступного визначення.

Визначення 2: Великі дані (англ. Big Data) в інформаційних технологіях – серія підходів, інструментів та методів обробки структурованих та неструктурованих даних великих обсягів і різноманітності для отримання результатів, які:

- 1) легко сприймаються людиною,
- 2) ефективні в умовах неперервного приросту, розподілення по численним вузлам обчислювальної мережі. [12]

Тобто "Big Data" – це не стільки обсяги даних, а перш за все технології та сукупність інструментів, які дозволяють розв'язувати задачі зберігання та обробки великих обсягів даних і дозволяють отримати результати у зрозумілій формі значно ефективніше ніж традиційні підходи.

До серії інструментів обробки "Big Data" належать системи масово-паралельної обробки невизначено структурованих даних – перш за все рішення категорії noSQL, алгоритми Map Reduce, програмні комплекси та бібліотеки Hadoop. [13]

В якості характеристик, які визначають поняття "Big Data", відзначають правило «три V»:

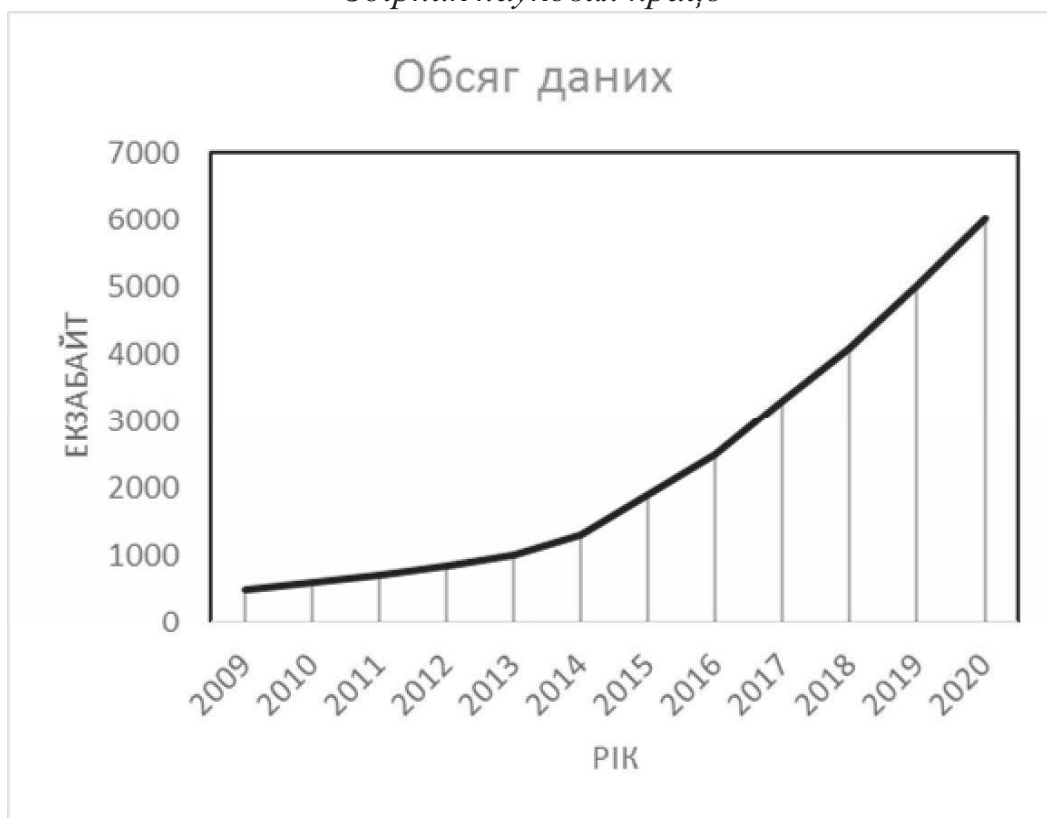
- 1) Volume – об'єм;
- 2) Velocity – швидкість, як у розумінні швидкості приросту, так і необхідності швидкої обробки та отримання результату;
- 3) Variety – різноманітність, у розумінні можливості одночасної обробки різних типів даних [12].

Таким чином, оскільки інструменти паралельної обробки Map Reduce та Hadoop створені для роботи з "Big Data" вони не можуть бути класифіковані інакше, незважаючи на те, що вони можуть бути застосовані до структурованих традиційних даних. Окремо обсяги даних також не впливають на відношення технології до "Big Data" оскільки очевидно, якщо технологія обробляє великі обсяги даних, то зможе обробити і невеликі. І з іншого боку, сучасні традиційні бази даних обробляють великі обсяги інформації, головна умова отримання результату – структурованість даних.

Можемо зробити висновок, що визначення "Big Data" – являється методологічною класифікацією підходів та інструментів для зберігання, обробки та аналізу даних, а не класифікацією самих даних. Тобто розміри даних в байтах та їх структурованість не є ознаками належності до "Big Data".

Розглянемо далі джерела формування неструктурованих даних великих обсягів. А також проблеми, пов'язані із швидким зростанням даних.

У період всього розвитку комп'ютерних технологій спостерігається прискорення приросту обсягів даних. Але останнім швидкість приросту даних зростає експоненціально (рис. 1).



Джерело: Розробка автора

Рис. 1. Обсяги даних та прогноз на 2020 рік.

В якості прикладів джерел виникнення великих даних наводяться [14] дані, що надходять з вимірювальних та побутових пристроїв, події від радіочастотних ідентифікаторів, потоки повідомлень із соціальних мереж, метеорологічні дані, дані дистанційного зондування Землі, потоки даних про локацію абонентів мереж мобільного зв'язку, пристроїв аудіо- і відеореєстрації. Очікується, що розвиток і початок широкого використання цих джерел даних ініціює проникнення технологій "Big Data" як в науково-дослідну діяльність, так і в комерційний сектор і сферу державного управління.

На сьогоднішній день у мережах роздрібній торгівлі, наприклад Walmart, обробляють більше 1 млн транзакцій клієнтів щогодини, які вводяться в бази даних та містять більш 2,5 петабайт (2560 терабайт) даних, або еквівалент

інформації, що міститься у всіх книгах в Бібліотеці Конгресу США збільшений у 167 разів [15].

У секторі нерухомості деякі компанії використовують анонімні сигнали GPS від майже 100 000 000 водіїв, щоб допомогти новим покупцям будинку визначити їх час витрачений на дорогу на роботу та з роботи в залежності від різних годин дня [16].

В галузі науки при проведенні експериментів на адронному коллайдері встановлюють близько 150 млн датчиків, які фіксують дані 40 мільйонів разів на секунду [17]. В результаті, менш ніж 0,001% від потоку даних із датчиків усіх чотирьох експериментів на адронному колайдері складає 25 петабайт даних до реплікації і майже 200 петабайт після реплікації.

Значну частину інформації створюють також роботи, при взаємодії один з одним та іншими мережами – наприклад, сенсори та інтелектуальні пристрої. При таких темпах приросту даних, зростає потреба в ефективному використанні та отриманні корисної інформації від цих даних. І така потреба породжує нетривіальні для вирішення практичні задачі або, як їх називають проблеми "Big Data".

В якості проблем "Big Data" виділяють три типи задач:

1. зберігання та управління даними;
2. обробка неструктурованих даних;
3. аналіз великих обсягів даних.

Проаналізуємо кожен із пунктів окремо. Обсяги в сотні терабайт (10^{12} байт) не можуть бути легко організовані для зберігання та управління традиційними реляційними базами даних. В якості системи "Big Data" для зберігання та обробки даних використовують розподілені файлові системи. Однією із найбільш відомих

та поширених розподілених файлових систем для зберігання даних є система Hadoop.

Проблема неструктурованих даних зумовлена тим, що більша частина всієї інформації є неструктурованою - наприклад, тексти, зображення, аудіо та відео. В таких умовах для отримання корисної інформації потрібна глибока обробка великих обсягів даних. Для обробки неструктурованих даних використовують високопродуктивні алгоритми які аналізують певний обсяг інформації виділяючи з нею ключові, найбільш цінні для подальшого аналізу дані.

Про аналіз великих обсягів даних в літературі написано значно менше ніж про зберігання та управління. В основі алгоритмів аналізу покладено принципи Map-Reduce які представляють собою модель для розподілених паралельних обчислень. Принципи його роботи полягає в тому, що дані розподіляються на кілька робочих вузлів файлової системи для попередньої обробки - фаза Map, а потім агрегуються вже попередньо оброблені дані - фаза Reduce. Таким чином для обрахунку кінцевої суми алгоритм паралельно обчислює кілька проміжних результатів в кожному окремому вузлі даних, а потім обчислює загальний результат. Даний підхід довів свою ефективність та високу швидкість обрахунків на великих обсягах даних у порівнянні із традиційними підходами.

Також у звіті McKinsey [18] були виділені наступні методи аналізу "Big Data":

1) методи класу Data Mining:

- навчання асоціативним правилам (Association rule learning),
- класифікація (методи категоризації нових даних на основі принципів, які раніше застосовувались до вже відомих даних),

- кластерний аналіз,
- регресійний аналіз;

2) краудсорсінг – категоризація і збагачення даних за допомогою широкого, невизначеного кола осіб, залучених на підставі публічної оферти, без вступу в трудові відносини;

3) змішування і інтеграція даних (Data fusion and integration) – набір технік, що дозволяють інтегрувати різноманітні дані з різноманітних джерел для можливості глибинного аналізу, в якості прикладів таких технік, складових цей клас методів наводяться цифрова обробка сигналів та обробка природної мови (включаючи тональний аналіз);

4) машинне навчання, включаючи навчання з учителем і без учителя, а також Ensemble learning – використання моделей, побудованих на основі статистичного аналізу або машинного навчання для отримання комплексних прогнозів;

5) штучні нейронні мережі, мережевий аналіз, оптимізація, в тому числі генетичні алгоритми;

6) розпізнавання образів;

7) імітаційне моделювання;

8) просторовий аналіз (Spatial analysis) – клас методів, що використовують топологічну, геометричну і географічну інформацію;

9) статистичний аналіз, А/В-тестування та аналіз часових рядів;

10) візуалізація аналітичних даних – подання інформації у вигляді малюнків, діаграм, з використанням інтерактивних можливостей та анімації як для отримання результатів, так і для використання в якості вихідних даних для подальшого аналізу.

Описані вищезгадані методи та підходи до аналізу даних – це тільки невелика частина інструментів, які дозволяють аналізувати великі обсяги даних. Існують і інші способи, наприклад, застосування спеціальних масштабованих алгоритмів, ієрархічних моделей, навчання вікнами та інше. Аналіз великих баз даних – це нетривіальне завдання, яке в більшості випадків не вирішується універсальним алгоритмом, проте сучасні бази даних та аналітичні платформи пропонують достатній набір методів вирішення цієї задачі. При розумному їх застосуванні системи здатні переробляти терабайти даних з прийнятною швидкістю.

Розвинені країни все частіше використовують технології роботи з великими обсягами даних. Вже зараз є 4,6 мільярда абонентів мобільного зв'язку по всьому світу і від 1 до 2 млрд людей з доступом до Інтернету [19]. У період з 1990 по 2005 рік, більше 1 мільярда людей у всьому світі були класифіковані як середній клас, який означає зростання рівня грамотності, що у свою чергу призводить до зростання інформації. Зростають технічні можливості для обміну інформацією через телекомунікаційні мережі – у 1986 році розмір трафіку становив 281 петабайт, у 1993 році – 471 петабайт, у 2000 році – 2,2 екзабайт, у 2007 році – 65 екзабайт і 667 екзабайт у 2014 році [15]. За однією з оцінок, одна третина глобально збереженої інформації є даними буквено-цифрового тексту і графічного формату. Такі формати являються найбільш корисними для аналізу методами "Big Data". Це також вказує на потенціал для застосування та розвитку методів аналізу відео та аудіо контенту.

Інтерес до теми "Big Data" зростає пропорційно до росту обсягів даних та розвитку самої технології. Останні два роки спостерігається значне зростання кількості

пошукових запитів у мережі Інтернет за цією темою (рис 2.) Слід також відмітити, що найбільшу зацікавленість до "Big Data" технологій проявляють в індійському регіоні та південно-східній Азії, а потім вже у США та ЄС.



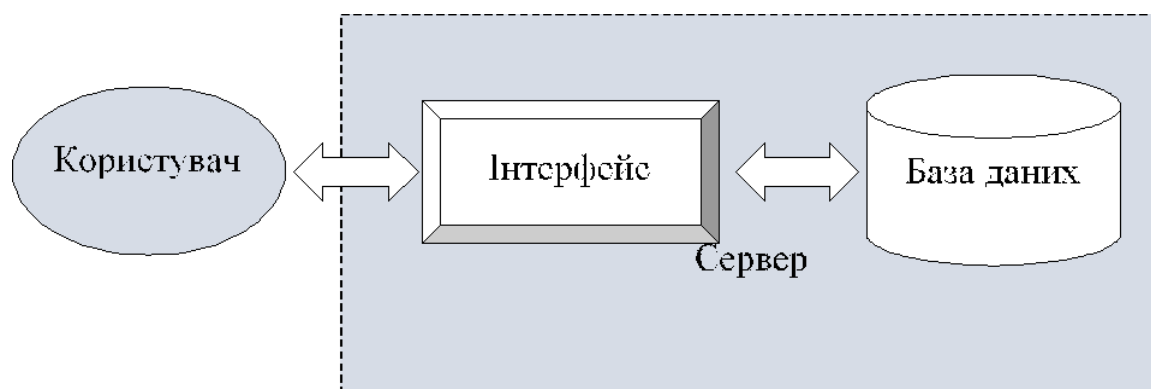
Джерело: Розробка автора

Рис. 2. Динаміка запитів "Big Data" у пошуковій системі Google.

Розглянемо далі на чому базуються технології "Big Data" та яка принципова відмінність від традиційних інструментів, а саме – модель роботи розподілених файлових систем, систему зберігання та обробки даних Hadoop та модель роботи алгоритмів Map Reduce.

Модель роботи традиційної системи управління базами даних передбачає наявність єдиного сервера для зберігання та обробки даних, програмного забезпечення – інтерфейсу, на основі якого користувачі можуть писати запити до бази даних та отримувати результати.

Схематично модель роботи традиційної СУБД зображена на рис. 3.



Джерело: Розробка автора

Рис. 3. Модель традиційної бази даних.

Така система добре працює з невеликими обсягами даних. Очевидним обмеженням є потужність одного сервера на якому розміщена база даних. Недоліками такої моделі є те, що на зберігання даних йде частина ресурсів сервера і коли даних стає досить багато, то на їх обробку витрачається дуже багато часу.

"Big Data" розв'язують дану проблему за допомогою алгоритмів паралельних обчислень Map Reduce.

Визначення 3: Map Reduce – модель розподілених обчислень, представлена компанією Google, яка використовується для паралельних обчислень над дуже великими (декілька петабайт) наборами даних у комп'ютерних кластерах [19].

Робота Map Reduce складається з трьох кроків: Map, Shuffle та Reduce. На Map кроці відбувається попередня обробка вхідних даних. Для цього один з комп'ютерів (його називають головним вузлом –master node) отримує вхідні дані завдання, поділяє їх на частини і передає іншим комп'ютерам (робочим вузлам –worker node) для

попередньої обробки. Назву даний крок отримав від однойменної функції вищого порядку.

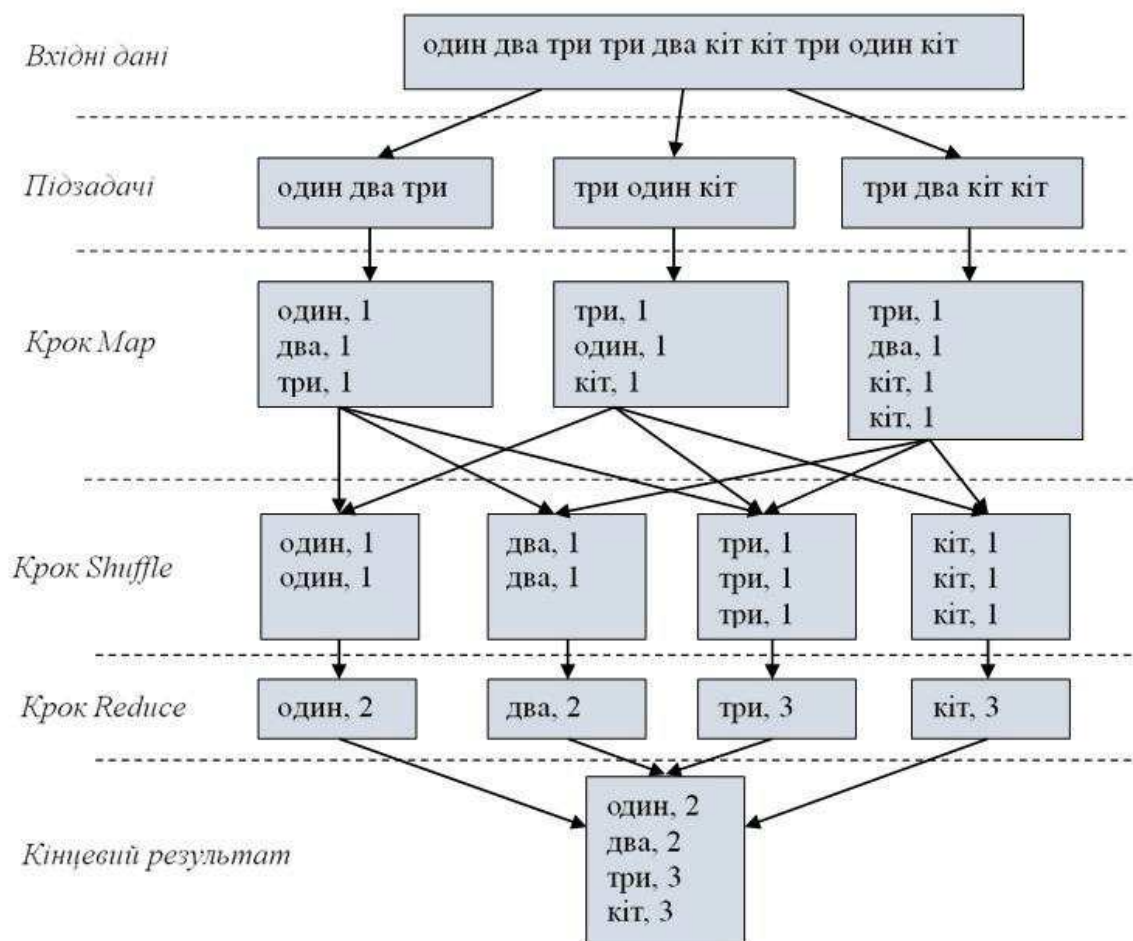
Крок Shuffle розділяє по групам результат роботи кроку Map. Кожна така група відповідає одному параметру. Далі ці групи використовуються як вхідні дані для функції Reduce.

На кроці Reduce відбувається агрегування попередньо оброблених даних. Головний вузол отримує відповіді від робочих вузлів і на їх основі формує результат – рішення задачі, яка формулювалася спочатку.

Перевага Map Reduce полягає в тому, що він дозволяє розподілено робити операції попередньої обробки і згортки. Операції попередньої обробки працюють незалежно одна від одної та можуть проводитися паралельно (хоча на практиці вони обмежені джерелом вхідних даних та/або кількістю робочих процесорів). Аналогічно, множина робочих вузлів може здійснювати згортку – для цього необхідно тільки щоб всі результати попередньої обробки з одним конкретним значенням ключа оброблялися одним робочим вузлом в один момент часу.

Серед задач які ефективно розв'язуються за допомогою підходу Map Reduce виділяють аналіз неструктурованого тексту. Наприклад, потрібно прочитати певний текст та підрахувати частоту кожного слова у цьому тексті.

Схематично робота алгоритму із вказаними кроками представлено на рис. 4.



Джерело: Розробка автора

Рис. 4. Модель роботи Map Reduce.

Рішення цієї задачі за допомогою мови програмування Python буде наступним:

1) Фаза Map:

```
def map(text):  
    text_map = []  
    for word in text.split():  
        text_map.append([word, 1])  
    return text_map
```

2) Фаза Shuffle

```
def shuffle(text_map):  
    text_shuffle = {}
```



```
for key in text_map:  
if key[0] not in text_shuffle.keys():  
text_shuffle[key[0]] = [1]  
else:  
text_shuffle[key[0]].append(1)  
return text_shuffle
```

3) Фаза Reduce

```
def reduce(text_shuffle):  
text_reduce = {}  
for key in text_shuffle.keys():  
text_reduce[key] = sum(text_shuffle[key])  
return text_reduce
```

В результаті почергового виконання вказаних функцій ми отримаємо результат у вигляді відповідних пар значень – слово та кількість згадувань у тексті. Такі розподілені обчислення можуть виконуватись на різних комп'ютерах для прискорення обрахунків та лише передавати результати для кінцевих обчислень на комп'ютер користувача. В результаті такого підходу швидкість виконання процедури значно зростає.

Висновки. Отже, в статті розглянуто питання що розуміти під терміном "Big Data", які його характерні властивості та ознаки, визначено джерела формування даних, виділено клас основних задач "Big Data" та розглянуто основні інструменти для розв'язання поставлених задач.

У статті також досліджені недоліки тредиційних систем управління базами даних та способи їх подолання за допомогою розподілених обчислень, що є основою роботи усіх методів "Big Data". В результаті розглянуто

модель процесу Map Reduce, побудовано діаграму процесу та програмну реалізацію на мові Python.

Можемо зробити висновок, що методи "Big Data" можна використовувати не тільки для даних великих розмірностей, але в такому випадку ефект від їх використання буде незначним. Тобто технології "Big Data" покликані розв'язати нові типи проблем, які виникли у зв'язку із різким збільшенням обсягів даних через доповнення класичних методів, а не їх заміну.

Література

1. Clifford Lynch, Big data: science in the petabyte era. Nature 2008, 455. P 1-50.
2. Майер-Шенбергер Виктор. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим/ Виктор Майер-Шенбергер, Кеннет Кукьер; пер. с англ. Инны Гайдюк. – М.: Манн, Иванови Фербер, 2014. – 240 с.
3. Dijcks Jean-Pierre. Big Data for the Enterprise // Oracle.– October, 2011. [Електронний ресурс] – Режим доступу: <http://BigDatawithoracle-521307.pdf>
4. Дубова Наталья. Большая конференция о Больших Данных(рус.). Открытые системы (3 ноября 2011). [Електронний ресурс] – Режим доступу: <http://www.osp.ru/news/articles/2011/45/13011213>
5. Finley Klint. Steve Ballmer on Microsoft's Big Data Future and More in This Week's Business Intelligence Roundup(англ.). Read Write Web (17 July 2011). [Електронний ресурс] – Режим доступу: <http://readwrite.com/2011/06/17/this-week-in-business-intellig>
6. Шах Агам. HP меняет персональные компьютеры на Большие Данные. Открытые системы (19 августа 2011). [Електронний ресурс] – Режим доступу: <http://www.osp.ru/news/articles/2011/34/13009952>
7. Н.Б. Шаховська. Модель Великих Даних «Сутність - характеристика». / Н.Б. Шаховська, Ю.Я. Болюбаш / 2015 р. [Електронний ресурс] – Режим доступу: http://www.academia.edu/19609620/%D0%9C%D0%9E%D0%94%D0%95%D0%9B%D0%AC_%D0%92%D0%95%D0%9B%D0%98%D0%9A%D0%98%D0%A5_%D0%94%D0%90%D0%9D%D0%98

[%D0%A5%D0%A1%D0%A3%D0%A2%D0%9D%D0%86%D0%A1%D0%A2%D0%AC-%D0%A5%D0%90%D0%A0%D0%90%D0%9A%D0%A2%D0%95%D0%A0%D0%98%D0%A1%D0%A2%D0%98%D0%9A%D0%90](#)

8. Черняк Леонид. Большие Данные – новая теория и практика // Открытые системы. СУБД. – М.: Открытые системы, 2011. – № 10. [Электронный ресурс] – Режим доступа: <http://www.osp.ru/os/2011/10/13010990/>
9. Ускенбаева, Р.К. Задачи создания больших данных - Big Data / Р.К. Ускенбаева, Ж.Б. Кальпеева, А. Касымова // Вычислительные и информационные технологии в науке, технике и образовании: международная конференция. - Усть-Каменогорск: ВКГТУ им.Д. Серикбаева, 2013. С 209 - 213.
10. Uskenbaeva, R.K. Tasks of resources provision of distributed computer system's functionality / R.K. Uskenbayeva, A.A. Kuandykov, A.U. Kalizhanova. – Dubai, World Academy of Science, Engineering and Technology.- 2012.- Iss.70. – P. 580-581.
11. Big Data – Wikipedia. [Электронный ресурс] – Режим доступа: https://en.wikipedia.org/wiki/Big_data
12. Большие данные – Wikipedia. [Электронный ресурс] – Режим доступа: https://ru.wikipedia.org/wiki/%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5
13. Моррисон Алан и др. Большие Данные: как извлечь из них информацию. Технологический прогноз. Ежеквартальный журнал, российское издание, 2010 выпуск 3. Pricewaterhouse Coopers. [Электронный ресурс] – режим доступа: <http://www.webcitation.org/6655TOKVD>
14. Канаракус Крис. Машина Больших Данных. Сети, № 04, 2011. Открытые системы. [Электронный ресурс] – режим доступа: <http://www.osp.ru/nets/2011/04/13010802/>
15. "Data, data everywhere". The Economist. 25 February 2010. [Электронный ресурс] – режим доступа: <chrome-extension://ecnphlgnajanjnkcmbrancdjoidceilk/http://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf>
16. Wingfield Nick. "Predicting Commutes More Accurately for Would-Be Home Buyers". [Электронный ресурс] – режим доступа:

http://bits.blogs.nytimes.com/2013/03/12/predicting-commutes-more-accurately-for-would-be-home-buyers/?_r=1

17. "LHC Guide, English version. A collection of facts and figures about the Large Hadron Collider (LHC) in the form of questions and answers.". CERN-Brochure-2008-001-Eng. CERN Document Server.
– [Електронний ресурс] – режим доступу:
<http://cds.cern.ch/record/1092437?ln=en>
18. Manyika James and others. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, June, 2011.
– [Електронний ресурс] – режим доступу:
<http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>
19. J. Dean and S. Ghemawat. Map Reduce: Simplified Data Processing on Large Clusters. In OSDI'04, 6th Symposium on Operating Systems Design and Implementation, Sponsored by USENIX, in cooperation with ACM SIGOPS, 2004. P. 137–150.

УДК 004:657

Л.В. Резниченко, О.О. Тімашов

Використання інформаційних технологій для малого бізнесу

Розглядається можливість застосування web-інтеграції та бухгалтерії в Excel як альтернативи ІС Бухгалтерії для синхронізованого функціонування сайту, складу, бухгалтерії та інших спеціалізованих програм для підприємств з невеликими щорічними оборотами на прикладі ФОП та підприємств малого бізнесу з метою ефективної бізнес-взаємодії ресурсів підприємства або інтернет-магазинів з локальними інформаційними системами.

Ключові слова: ERP система, інтернет-магазин, інформаційна технологія, web-сайт, web-інтеграція.