

Технічні науки

УДК 05.13.01

**Тимошин Юрій Афанасійович**

Кандидат технічних наук, доцент, доцент кафедри технічної кібернетики  
Національний технічний університет України «Київський політехнічний  
інститут»

**Гохкаленко Сергій Дмитрович**

студент  
Національний технічний університет України «Київський політехнічний  
інститут»

**Тимошин Юрий Афанасьевич**

Кандидат технических наук, доцент, доцент кафедры технической кибернетики  
Национальный технический университет Украины "Киевский  
политехнический институт"

**Гохкаленко Сергей Дмитриевич**

студент  
Национальный технический университет Украины "Киевский  
политехнический институт"

**Timoshyn Y.**

candidate of engineering sciences, Associate Professor at the Department of  
technical cybernetics

National Technical University of Ukraine "Kyiv Polytechnic Institute"

**Hokhkalenko S.**

student National Technical University of Ukraine "Kyiv Polytechnic Institute"

## **ЦІЛЬОВЕ ЗАСТОСУВАННЯ ТА ОСНОВНІ ТЕНДЕНЦІЇ СУМІСНОГО ВПРОВАДЖЕННЯ BIG DATA ТА IOT**

## **ЦЕЛЕВОЕ ПРИМЕНЕНИЕ И ОСНОВНЫЕ ТЕНДЕНЦИИ СОВМЕСТНОГО ВНЕДРЕНИЯ BIG DATA И IOT**

## **INTENDED USE AND MAJOR TRENDS OF THE JOINT IMPLEMENTATION OF BIG DATA AND IOT**

**Анотація:** Досліджені області застосування та основні реалізації технологій Big Data та Internet of Things, їх переваги.

**Ключові слова:** Big Data, Internet of Things, Nadoor, датчики.

**Аннотация:** Исследованы области применения и основные реализации технологий Big Data и Internet of Things , их преимущества.

**Ключевые слова:** Big Data, Internet of Things, Hadoop, датчики.

**Summary:** Investigated the scope and basic implementation of Big Data technologies and Internet of Things, their benefits.

**Key words:** Big Data, Internet of Things, Hadoop, sensors.

Інтернет речей (IoT) і великі дані – дві великі і актуальні теми які стосуються розробки комерційних, промислових та інших додатків.

Термін "IoT" був введений в 1999 році і відноситься до світу пристроїв, підключених до Інтернету[1].

Основним призначенням IoT і Big Data є збір і аналіз даних.

В сучасному світі цей функціонал є дуже потрібним для бізнесу. Він дозволяє проаналізувати поведінку покупців [2], зокрема що вони купляють і чому.

Прикладом цього можуть служити картки лояльності, які покупці використовують в продуктових магазинах та інших торгових точках. За допомогою цих карт, роздрібні продавці та їх постачальники можуть з'ясувати, які клієнти купують, які продукти, а потім можуть використовувати цю функцію для збільшення продажів і прибутку.

Мета цієї статті - показати, чому виробники повинні інтегрувати IoT і великі дані в свої системи промислової автоматизації, а також про те як це можна зробити, використовуючи продукти і системи доступні сьогодні. Метою роботи є аналіз технологій для роботи з великими даними та IoT, а також пошук оптимального рішення цієї проблеми. При роботі з великими даними потрібно розглянути різні технології, які зараз можна використати

для вирішення такої проблеми. В нас час з кожними роком розмір даних тільки збільшується. Потрібно порівняти наявні технології та рішення. Зробити висновки про перспективи та напрям розвитку.

В статті розглянуті перспективи сумісного використання та майбутнє розвитку IoT і великих обсягів даних.

### **Чому потрібно впроваджувати IoT і Big Data?**

Як Sanat Joshi пояснює в InTech ISA, в "Big Data є новою нормою для корпоративної аналітики і є досить поширеною у багатьох галузях промисловості".

Він каже, що дані стають «Big Data» коли обсяг, швидкість і / або

різноманітність даних перевищують можливості поточної IT-систем, приймати, зберігати, аналізувати або іншим чином використовувати її.

Збір великої кількості даних[3] почався з обмежених вимірювань людини на основі спостережень, і інформації, яка вводиться в систему людиною. Більшість даних в наш час автоматично генерується за допомогою датчиків та сенсорів, які в свою чергу стають все більш чисельними і розумнішими та з ширшим функціоналом.

Розвиток IoT дозволяє збільшити можливості звітності для зовнішніх пристроїв та додавати нові данні в залежності від їх типу (табл. 1) про об'єкт спостереження[4].

1. Температура
2. Тиск
3. Рівень
4. Потік
5. Щільність
6. Близькість
7. Положення , кут , зсув , відстань , швидкість , прискорення
8. Бачення
9. Вібрація
10. Струм , напруга та інші параметри живлення
11. рН та інші аналітичні значення
12. Масова частка вологи , вологості та інших погодних умов

Табл. 1. Типи датчиків



Рис. 1. Области використання Big Data та IoT

На рис. 2 приведені інвестиції в Big Data в конкретних галузях виробництва де ефективно використовуються ці технології. Але їх використання не обмежується самим виробництвом та управлінням технологічними процесами напрому. Активне застосування знайшлося і при упаковці, а також виконання випробувань і перевірок.

Нижче (табл. 2) наведені приклади задачі і цілі які можуть бути здійснені впровадженням IoT і Big Data. Їх використання приведе до підвищення продуктивності, пропускну здатності, скорочення часу простоїв, а також зниження витрат на додаткове обладнання.

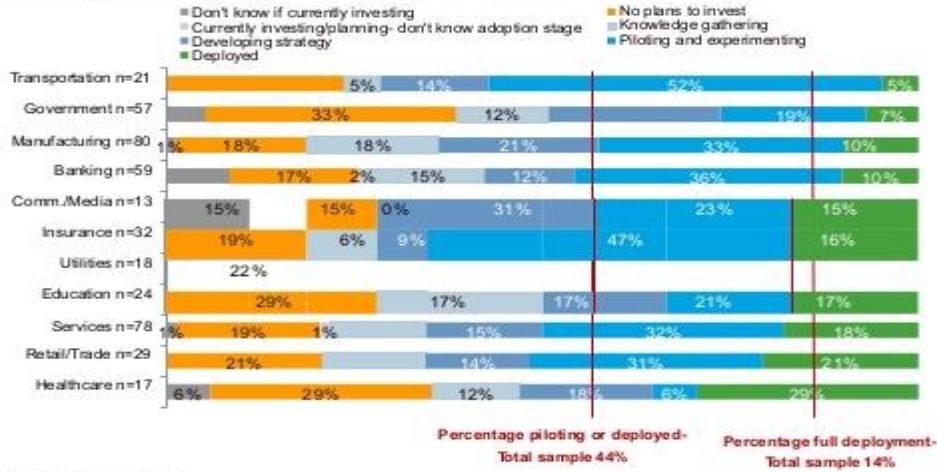
Також на рис. 2 наведені процентні співвідношення використання Big Data по галузях в світі [5].

1. Інтелектуальний моніторинг стану здоров'я
2. Зменшення часу простоїв
3. Нижній відсоток браку
4. Підвищення якості
5. Вища пропускна здатність
6. Підвищення рівня безпеки
7. Ефективне використання робочої сили
8. Включити масову настройку

Таб. 2. Конкретні існуючі приклади використання Big Data та IoT разом

## State of Big Data adoption- by industry

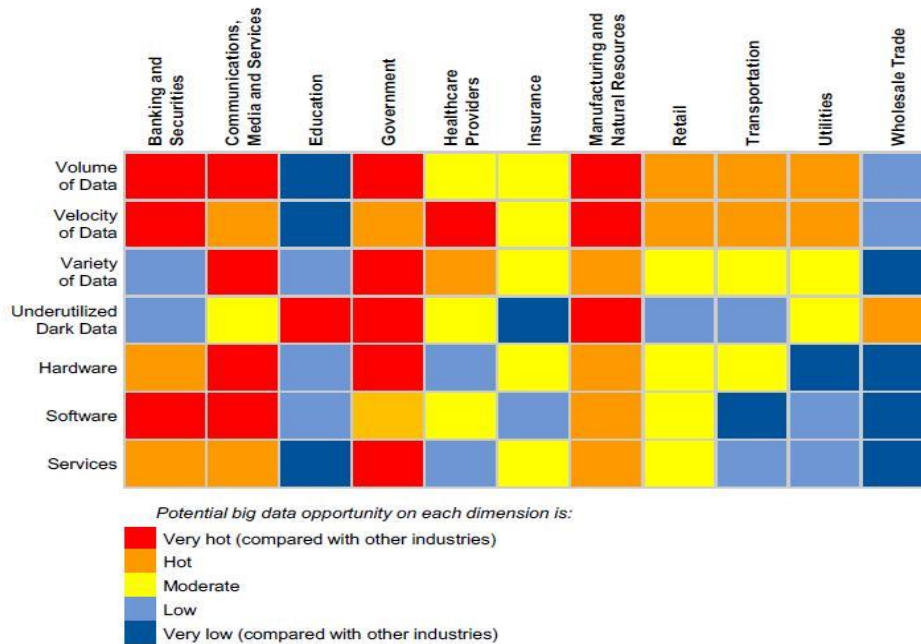
Which of the following best describes your organization's stage of big data adoption?



© 2015 Gartner, Inc. All rights reserved.

Рис. 2. Процентне впровадження Big Data по галузях в світі на 2015 рік.

Figure 2. Big Data Opportunity Heat Map by Industry



Source: Gartner (July 2012)

Рис. 3. Можливості впровадження Big data по галузям[5].

Оскільки тема Big Data проходить через етап стрімкого розвитку зараз, тому існує багато різноманітних концепцій, реалізацій і підходів до вирішення цієї задачі. Нижче описані основні тенденції в Big Data на 2016 рік.

## 1. NoSQL поглинання

NoSQL технології, як правило, пов'язані з неструктурованими даними, за останній рік зробили великий крок в області Big Data. Забігаючи вперед, можна зауважити, що міграція до баз даних NoSQL в якості основних для IT-архітектури підприємства показує, як переваги концепцій схеми бази даних стають більш вираженими.

Gartner Magic Quadrant для компаній операційних баз даних показує значне зменшення долі на ринку систем управління, де в минулому домінували Oracle, IBM, Microsoft і SAP[6].

На відміну від цього, в самому останньому Magic Quadrant, компанії NoSQL, включаючи MongoDB, DataStax, Redis Labs і MarkLogic чисельно перевершують традиційні бази даних постачальників і розташовані в квадранті лідерів, визначених Gartner.

## 2. Apache Spark «просвітлює» Big Data

Apache Spark вперше появився будучи складовою частиною екосистеми Hadoop на великій платформі даних вибору для цілого ряду підприємств.

Apache Spark забезпечує значно більшу швидкість обробки даних в порівнянні з Hadoop і в даний час проект з відкритим вихідним кодом для найбільших об'ємів Big Data [7].

Все більше і більше з'являється успішних прикладів використання інфраструктури навколо Spark, наприклад, в Goldman Sachs, де Spark став ведучою технологією серед усіх впроваджених технологій для аналітики [8].

3. Проекти Hadoop: підприємства продовжують свій рух від Hadoop концепції до виробництва

В ході недавнього опитування із 2200 клієнтів Hadoop, всього лише 3% респондентів висловили очікування, що вони будуть розробляти менше з Hadoop протягом наступних 12 місяців, і 76% тих, хто вже використовував Hadoop, планував застосовувати більше протягом найближчих трьох місяців.

4. Big Data зростає: Hadoop стає корпоративним стандартом

Ще одним свідченням зростаючої тенденції, що Hadoop стає одною із основних компонент ІТ-інфраструктури підприємства, означає ріст інвестицій в компонентах, які оточують корпоративні системи, такі як безпека.

Проект Apache Sentry забезпечує систему для забезпечення дотримання доступу, на основі ролей дозволу на доступ до даних і метаданих, що зберігаються на кластері Hadoop[9].



В даний час на перший план виходять нові високі технології передачі даних, які тим самим усувають ще один бар'єр прийняття Big Data на підприємствах у вигляді покращання безпеки та швидкостей збору інформації, вартості обладнання.

5. Значне покращення показників швидкості в Big Data: варіанти розширення, щоб реалізувати зростання швидкості аналізу в Hadoop.

Hadoop набере більше вагомості на підприємстві, якщо буде зростаючий попит з боку кінцевих користувачів, які отримують такі ж можливості дослідження швидкості даних, які вони звикли очікувати від традиційних сховищ даних.

Hadoop задовільнить попит кінцевих користувачів на впровадження таких технологій, як Cloudera Impala, AtScale, Actian Vector і Jethro. Технологія Hadoop буде рости далі, чим зробить нечіткою межу, яка розділяє основи "традиційної" ВІ концепції і світ Big Data.

6. З'являються інструменти для самостійного обслуговування користувачів, які дозволяють керувати процесом аналітики та відображення інформації на основі вже зібраних даних. Це зумовлено стрімким розвитком і орієнтованістю на бізнес та кінцевого користувача.

7. Зростання MPP(Massively parallel processing) сховищ даних. Хмарні технології набувають популярності

Помилки та навіть механічні поломки сховищ даних, які вели до повної або часткової їх втрат, було важливим питанням протягом деякого часу. Проте це не секрет, що зростання в цьому сегменті ринку сповільнюється.

Але зараз існує значний зсув в застосуванні хмарних технологій, де Amazon реалізував хмарні ресурси на вимогу в Redshift.

Redshift - швидко зростаючий сервіс AWS, проте зараз з'явилися серйозні конкуренти з боку Google з продуктом BigQuery, пропозиції щодо сховищ даних від великих компаній, таких як Microsoft (з Azure SQL сховищами даних) і Teradata. Також конкуренцію їм складають нові стартапи, такі як Snowflake, (переможець Strata + Hadoop World 2015) який також набуває популярності в цій галузі.

Аналітики цитують 90% компаній, які взяли за основу Hadoop і будуть використовувати його з хмарними технологіями. Ці клієнти можуть динамічно масштабувати обсяг пам'яті і обчислювальних ресурсів які використовуються для збереження великих обсягів інформації, що зберігається в їх Hadoop екосистемі

## 8. Ефективне використання: IoT, хмара і великі дані об'єднуються

Хоча ця технологія відносно нова, але дані від пристроїв з Інтернету речей (IoT) стають одними з основних "вбивців додатків" для хмари і як генератор великого масштабу кількості інформації вимірюваної в петабайтах.

Великі компанії, такі як Google, Amazon Web Services і Microsoft, поступово переходять на хмарні технології та проводять міграцію всіх

даних в хмарі. Тому в недалекому майбутньому вся інформація буде плавно переходити до їх хмарних аналітичних систем.



Рис. 4. Магічний квадрат Gartner з найбільшими компаніями-вендорами Big Data

На рис. 4 бачимо основні компанії які займаються впровадженням розглянутих в статті технологій на ринку та їх взаєморозташування[10]. Серед лідерів можна виділити Teradata, Oracle, IBM –компанії які мають відповідно свої рішення, що користуються найбільшим попитом у світі. З магічного квадрату Gartner також бачимо, що така велика компанія, як Microsoft, яка є абсолютним лідером в багатьох ІТ- областях є лише

претендентом. Big Data технології в Microsoft лише на етапі розвитку, хоча вже існує цілий ряд сервісів, такі як Microsoft Azure— хмарна технологія, яка буде сприяти стрімкому розвитку і об'єднаному використанню сукупності технологій, розглянутих в статті.

Велика кількість IT-науковців висловлюють думку, що пік Big Data, який почався декілька років тому, пройшов та інтерес спеціалістів до цієї області значно спав. Вони пояснюють це тим, що технології для рішення цих задач розвиваються швидше ніж області застосування. Вхідними даними для обробки є структуровані та неструктуровані данні, найчастіше логи програм та різні повідомлення які генеруються автоматично. Вони вважають ці дані мусором, який корисний у випадку некоректної роботи програми, виконання транзакцій та ін.

Автори рахують, що за цими технологіями майбутнє, тому що вони допомагають скласти загальну картину глобальних процесів таких як переміщення товарів, моніторинг систем з різноманітними типами датчиків та зробити їх більш ефективними. До того ж дані такого аналізу та результати успішно можна використати в області штучного інтелекту[11]. Для отримання даних для проведення аналітики немає кращого джерела ніж Internet of Things, які включають в собі різноманітні пристрої. Особливо актуальними є дані з мобільних пристроїв (смартфонів, планшетів і ін.), які включають в себе велику кількість датчиків та генерують інформацію про поведінку людини, її інтереси, звички та спосіб життя. Наприклад, програма моніторингу фізичної активності збирає дані про дистанцію пройденого шляху та кількості спалених калорій на протязі дня.



Рис. 5. Кількість створених даних та наявних ресурсів

Не стоїть на місці розвиток фізичних сховищ даних та хмарних технологій, що безумовно позитивно відображається на можливостях створення колекцій даних. Проте цей розвиток недостатній, про що свідчить співвідношення доступних ресурсів до кількості інформації (рис. 5). Це відповідно штовхає інженерів на виділення з цього великого об'єму корисної інформації, яка буде задіяна в процесі планування та прийняття рішень.

Резумуючи описане вище, можна стверджувати, що зв'язка технологій Big Data та IoT досі на піку свого розвитку. Ці технології стають доступними не лише для великих компаній, як це було спочатку, а стек технологій став нормою для проектів, які зможуть стрімко та успішно розвиватись.

## Література:

1. Kevin Ashton. That 'Internet of Things' Thing. In the real world, things matter more than ideas. (англ.). RFID Journal (22 June 2009). Проверено 30 ноября 2012. Архивировано из первоисточника 24 января 2013
2. Tamara Dull. .Big data and the Internet of Things: Two sides of the same SAS Best Practices [Электронный ресурс – Режим доступа: [http://www.sas.com/en\\_us/insights/articles/big-data/big-data-and-iot-two-sides-of-the-same-coin.html](http://www.sas.com/en_us/insights/articles/big-data/big-data-and-iot-two-sides-of-the-same-coin.html)]
3. Черняк, Леонид Большие Данные — новая теория и практика (рус.) // Открытые системы. СУБД. — М.: Открытые системы, 2011. — № 10. — ISSN 1028-7493.3. Моррисон, Алан и др. Большие Данные: как извлечь из них информацию (рус.). Технологический прогноз. Ежеквартальный журнал, российское издание, 2010 выпуск
4. Како Н., Яманэ Я. Датчики и микро-ЭВМ. Л: Энергоатомиз дат, 1986г.
5. Электронный ресурс: <http://www.forbes.com/sites/louiscolombus/2012/08/16/roundup-of-big-data-forecasts-and-market-estimates-2012/#548e58473381>
6. Электронный ресурс :<http://www.intersystems.com/our-products/cache/intersystems-recognized-leader-gartner-magic-quadrant-operational-dbms/>
7. Электронный ресурс: <https://spark.apache.org>
8. Zaharia, Matei; Chowdhury, Mosharaf; Franklin, Michael J.; Shenker, Scott; Stoica, Ion. Spark: Cluster Computing with Working Sets(PDF). USENIX Workshop on Hot Topics in Cloud Computing (HotCloud)
9. .Preimesberger, Chris Hadoop, Yahoo, 'Big Data' Brighten BI Future (англ.). EWeek (15 August 2011). Проверено 12 ноября 2011. Архивировано из первоисточника 17 мая 2012.
10. PricewaterhouseCoopers .Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data (англ.). Gartner (27 June 2011)
11. Manyika, James et al. Big data: The next frontier for innovation, competition, and productivity (англ.) (PDF). McKinsey Global Institute, June, 2011. McKinsey (9 August 2011).