

УДК 004.652.4:378.147

Ситник Ніна Василівна

кандидат економічних наук, доцент, професорка кафедри інформаційних систем в економіці
ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана», м. Київ, Україна
ORCID ID 0000-0002-4032-4881
vfsytnyk@gmail.com

Зінов'єва Ірина Сергіївна

кандидат економічних наук, доцентка кафедри інформаційних систем в економіці
ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана», м. Київ, Україна
ORCID ID 0000-0001-5122-8994
ira.zinovyyeva@kneu.edu.ua

СУЧАСНІ БАЗИ ДАНИХ NoSQL У ПІДГОТОВЦІ БАКАЛАВРІВ СПЕЦІАЛЬНОСТІ "КОМП'ЮТЕРНІ НАУКИ"

Анотація. Стаття присвячена проблемі опанування сучасними підходами організації баз даних майбутніми фахівцями зі спеціальності 122 «Комп'ютерні науки». Бази даних NoSQL і NewSQL – новітні тренди у сфері ІТ, які потребують детального ознайомлення та вивчення студентами комп'ютерних спеціальностей. У статті досліджено та розкрито поняття баз даних NoSQL і NewSQL. Проведена порівняльна характеристика реляційних баз даних з базами даних типу NoSQL, NewSQL. У зв'язку з поширенням різного виду WEBсервісів, потребою розподіленого оброблення великих масивів даних (Big Data) обґрунтовано необхідність повноцінного вивчення нового покоління баз даних та отримання фахових компетентностей щодо їх практичного застосування. Вивчення новітніх баз даних не має бути альтернативою вивченню реляційних баз даних. Враховуючи концептуальні відмінності і технологічні особливості використання сучасних баз даних, у Київському національному економічному університеті імені Вадима Гетьмана до навчальних планів спеціальності 122 «Комп'ютерні науки» додано нову дисципліну «Організація баз даних NoSQL», вивчення якої поповнить знання студентів про основні тенденції розвитку баз даних та підвищить їх конкурентоздатність на ринку праці. Під час розробки навчальної програми дисципліни були проаналізовані навчальні плани підготовки спеціалістів з комп'ютерних наук провідних закладів вищої освіти України, літературні та інтернет-джерела, які підтвердили необхідність та доцільність вивчення баз даних NoSQL та NewSQL. Дослідження критеріїв та аналіз рейтингу систем керування базами даних дозволили дійти висновку про доцільність введення до навчальної програми дисципліни «Організація баз даних NoSQL» тем з вивчення найбільш поширених у світовій практиці баз даних типу NoSQL (MongoDB, Redis, Cassandra) та NewSQL (Microsoft SQL Server 2017), яка має розширений функціонал і підтримує роботу з графами.

Ключові слова: реляційні бази даних; NoSQL; NewSQL; системи керування базами даних; комп'ютерні науки.

1. ВСТУП

Постановка проблеми. Обов'язковою складовою підготовки фахівців з комп'ютерних наук є вивчення баз даних, оскільки вони є необхідним елементом цифрових технологій, які потрібно розвивати і впроваджувати в усі сфери згідно «Концепції розвитку цифрової економіки та суспільства України на 2018-2020 роки [1].

Донедавна метою підготовки студентів з організації баз даних було вивчення і набуття фахових компетентностей щодо організації та роботи з реляційними базами даних. Безумовно, реляційні бази даних сьогодні є найбільш поширеними і використовуються на практиці понад сорок років (сама теорія реляційних баз даних була розроблена Едгаром Ф. Коддом ще в 1970 р.). Без реляційних баз даних складно

увести роботу сучасних інформаційних систем організаційно-економічного управління, зокрема в питаннях, що стосуються виробництва та промисловості, банківського і фінансового секторів, бухгалтерського та управлінського обліку, управління кадрами тощо. Іншими словами, реляційні бази даних використовуються всюди, де необхідним є опрацювання структурованих даних.

В останні десятиліття почали стрімко розвиватися нові напрями застосування інформаційних технологій, як-от: різного роду WEBсервіси, мобільні додатки, геоінформаційні системи, мультимедіа, Grid-технології, системи штучного інтелекту і т.п. Вони, у переважній більшості, вимагають оперування великими обсягами неструктурованих даних, підтримки одночасної роботи великої кількості користувачів, горизонтального масштабування, застосування хмарних технологій поряд із забезпеченням високого рівня продуктивності, відмовостійкості та надійності. Ці потреби неможливо було забезпечити засобами реляційних баз даних, що стало поштовхом для формування концепції та практичної розробки баз даних NoSQL, а згодом – баз даних NewSQL.

Зміни у світових тенденціях розвитку ІТ та баз даних зокрема вимагають постійного оновлення навчальних планів закладів вищої освіти (ЗВО), особливо для спеціальностей, що готують майбутніх фахівців у цій сфері.

Аналіз освітньо-професійних програм спеціальності 122 «Комп'ютерні науки» показав, що вивчення баз даних є обов'язковою дисципліною бакалаврського рівня підготовки. Вивчення дисципліни, яка в переважній більшості ЗВО називається «Організація баз даних та знань», присвячено реляційним базам даних.

У деяких ЗВО, наприклад, у Харківському національному університеті радіоелектроніки, до навчальних планів для спеціальності 122 «Комп'ютерні науки» додана дисципліна бакалаврського циклу «Спеціалізовані бази даних (NoSQL)», яка пропонує на теоретичному рівні вивчення класифікації NoSQL та сфери їх використання, практична частина курсу основана на роботі з документарноорієнтованою базою в середовищі системи керування базами даних MongoDB [2].

У Національному технічному університеті «Київський політехнічний інститут імені Ігоря Сікорського» для спеціальності 122 «Комп'ютерні науки» викладаються предмети з організації баз даних, роботи розподілених баз даних, а студентам магістерського рівня пропонується дисципліна «Постреляційні бази даних». З огляду на контекст викладеного на сайті університету освітнього контенту основна увага в підготовці фахівців з комп'ютерних наук приділяється опануванню MySQL, MS SQL SERVER, ORACLE, Caché, MongoDB та Cassandra [3], що свідчить про вивчення студентами сучасних типів баз даних.

У Київському національному економічному університеті ім. Вадима Гетьмана реляційні бази та сховища даних вивчаються в курсі «Організація баз та сховищ даних». До нового навчального плану підготовки бакалаврів від 2018/19 н.р. спеціальності 122 «Комп'ютерні науки» додана дисципліна «Організація баз даних NoSQL». З 2019/20 н.р. дана дисципліна буде нормативною, вивчатиметься у 5 семестрі з загальним обсягом 5 кредитів і міститиме: 18 годин лекційних, 14 годин практичних та 16 годин лабораторних занять, 15 годин відведено на індивідуальну роботу і 83 години на самостійну роботу. Мета вивчення дисципліни – сформулювати фундаментальні теоретичні знання та практичні вміння з організації нереляційних баз даних NoSQL та дати студентам уявлення про бази даних NewSQL. У рамках вивчення дисципліни майбутні фахівці з комп'ютерних наук мають опанувати сучасні технології проектування документарних, колонкоорієнтованих та графових баз даних (MongoDB,

Cassandra, Redis), а також набути професійних навичок роботи в середовищі сучасних систем керування базами даних NewSQL (на прикладі Microsoft SQL Server 2017) [4].

Однак більшість ЗВО, що здійснюють підготовку фахівців зі спеціальності 122 «Комп'ютерні науки» або не мають у своєму арсеналі дисциплін, які присвячені організації баз даних NoSQL та NewSQL, або під вивчення сучасних тенденцій розвитку баз даних відведено лише кілька годин у межах загальних курсів. На наш погляд, така ситуація є істотним недоліком у підготовці майбутніх фахівців з комп'ютерних наук, оскільки студенти повинні вивчати всі інноваційні напрями розвитку IT-технологій. Окрім того, вивчення організації баз даних NoSQL та NewSQL є теоретичним підґрунтям для дисципліни «Технології оброблення та аналіз надвеликих масивів даних», ряду дисциплін магістерського рівня та для написання кваліфікаційних бакалаврських (магістерських) робіт.

Аналіз останніх досліджень і публікацій. Домінуюча роль баз даних у реалізації сучасних IT підтримує на високому рівні науково-практичний інтерес до питань їх еволюції, розвитку, виявлення та усунення недоліків їх роботи шляхом постійного вдосконалення.

Загальні та методологічні аспекти проектування, розробки та супроводу різних типів баз даних представлені, переважно, у навчально-довідковій літературі [5], [6], [7], [8], значно менше – у наукових джерелах [9]. Більш детальну увагу на відстеженні сучасних тенденцій розвитку баз даних сфокусовано в періодичних джерелах [10], [11]. Більшість проаналізованих публікацій містять у собі елементи порівняння та протиставлення різних типів баз даних. Зокрема у публікаціях А. Абдулли (Ahmad Abdullah) та Ц. Чжуге (Qingfeng Zhuge) [12], В.О. Брацького, О.М. М'якишило [13], С.О. Спасітелевої, В.Л. Бурячка [14], В. Шарма (Vatika Sharma), М. Дейва (Meenu Dave) [15], М.Ю. Швеця, Д.С. Заруби, Ю.В. Хохлова [16] сфокусовано увагу на порівнянні відмінностей між реляційними та нереляційними базами даних. Згадані автори акцентують увагу на пошуку критеріїв співставлення SQL та NoSQL баз даних та визначенні методики оцінювання «кращого» (з позиції універсальності) типу баз даних.

Аналіз праць П. Аценіна (Paolo Atzenin), Ф. Бугіотті (Francesca Bugiotti), Л. Россі (Luca Rossi) [10], Ж.-Р. Лоурензоля (João Ricardo Lourenzol) та ін. [11] показав, що актуальним трендом у зарубіжній науковій думці є відхід від протиставлення баз даних SQL та NoSQL. У реальному секторі (різними бізнес-структурами) активно використовуються як одні, так і інші типи баз даних, просто кожен тип застосовується для вирішення специфічного кола задач. Більше того, сьогодні має місце розвиток нової концепції взаємодії SQL та NoSQL, яка отримала назву NewSQL. Бази даних NewSQL містять у собі сильні сторони реляційних баз даних (синхронність транзакцій, нормалізованість тощо) у поєднанні з підвищеною продуктивністю, горизонтальною масштабованістю. Публікації останніх років зосереджені на вивченні ролі та місця баз даних NoSQL і NewSQL у сучасній кон'юнктурі ринку IT. Науковцями С. Бінані (Sneha Binani), А. Гутті (Ajinkya Gutti), Ш. Ападхаєй (Shivam Upadhyay) [17] та Д. Кепнером (Jeremy Kepner) [18] проводяться паралелі між існуючими типами баз даних: SQL, NoSQL, NewSQL.

Проведений аналіз показав, що досить багато науковців з усього світу приділяють свою увагу вивченню, аналізу та прогнозуванню подальших тенденцій розвитку баз даних, що, безумовно, вказує на важливість та актуальність даного питання. Втім під час розробки курсу «Організація баз даних NoSQL» автори зіштовхнулись із проблемою вибору конкретних систем керування базами даних для наповнення практичної та лабораторної частин дисципліни. На сучасному ринку IT представлено більше 300 різноманітних баз даних та систем їх керування, із них тільки 1/3 є традиційними реляційними базами даних, інші ж – належать до типів NoSQL та

NewSQL [19]. Така велика кількість систем ставить проблему зваженого і обґрунтованого підходу до вибору тих систем керування базами даних, які мають вивчати студенти. Складність вибору полягає, на нашу думку, і у відсутності для викладачів методичних орієнтирів при формуванні навчальних програм власних курсів, що пов'язані з вивченням баз даних типу NoSQL і NewSQL.

Мета статті – дослідження світових тенденцій розвитку і системний огляд сучасних баз даних типу NoSQL і NewSQL для підвищення якості підготовки майбутніх фахівців з комп'ютерних наук.

2. ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Аналіз наукових досліджень та публікацій дав підставу виділити три ключові типи баз даних: реляційні або SQL, NoSQL (у значенні Not Only SQL) та NewSQL. Відзначимо, що в зарубіжних публікаціях останніх років [18], [20] згадується новий тип баз даних «мультисховище» (polystore databases), які являють собою сховища даних з інтегрованими реляційними та нереляційними моделями даних і, фактично, об'єднують у собі всі попередні типи баз даних. Враховуючи те, що наукова та практична база розробки цього типу баз даних тільки формується, у цій публікації буде зосереджено увагу на перших трьох типах (SQL, NoSQL, NewSQL).

Реляційні бази даних – це чітко структуровані дані, що характеризують певну предметну область і представляються у вигляді логічно пов'язаних поіменованих двовимірних плоских таблиць, що знаходяться під управлінням системи керування базами даних (СКБД) [6]. Основи реляційних моделей найбільш повно викладено в роботах Дж. Дейта [5], К.Т. Конноллі [21] і Дж. Мартіна [22]. Такі відомі представники класу SQL, як PostgreSQL, MySQL, Oracle активно використовуються з 80-х рр. XX ст. і дотепер.

NoSQL – це нове покоління баз даних, що характеризується моделями та технологіями функціонування, які відмінні від традиційних реляційних моделей. Вони виникли як реакція на існуючі проблеми функціонування традиційних баз даних. Вперше термін «NoSQL» був використаний Карло Строцці для бази даних, що не використовує мову запитів SQL, а в 2009 році Ерік Еванс і Йохан Оскарсон використали його для нереляційних баз даних [23]. З того часу термін NoSQL стосується баз даних, моделі яких відрізняються від реляційних (фактично є нереляційними). Характерними особливостями баз даних NoSQL є: відсутність мови запитів SQL та механізму підтримки транзакцій з виконанням ACID-вимог (Atomicity – атомарність, Consistency – узгодженість, Isolation – ізолюваність, Durability – довговічність зберігання); робота в розподіленому середовищі; горизонтальне масштабування; можливість опрацювати агреговані, неструктуровані, ненормалізовані великі обсяги даних; відсутність жорстко заданої схеми зберігання даних. Бази даних NoSQL підтримують BASE-вимоги (Basically Available – базова доступність, Soft state – гнучкий стан, Eventually consistent – узгодженість в підсумку) [7], [8], [24].

Описані характеристики є спільними для баз даних NoSQL, проте жодна з них не є визначальною. Наприклад, М. Фаулер і Дж. Садаладж Прамодкумар стверджують, що NoSQL – це випадковий неологізм, який поширився як спільна назва покоління баз даних, вектор розвитку яких протилежний реляційним системам [7]. Абревіатура NoSQL (Not Only SQL) перекладається як «не лише SQL». Не зважаючи на те, що термін є невдалим, він не заперечує реляційні бази даних та мову SQL, а лише вказує на те, що бази NoSQL мають право на існування і використання в тих предметних областях, де є проблематичним використання табличного, тобто реляційного представлення даних.

На даний момент немає чітко визначеної класифікації баз даних NoSQL. Більшість науковців сходиться у тому, що всі бази даних NoSQL можна умовно розділити на чотири групи (за моделлю організації даних): модель ключ/значення (Key/Value), таблична колонкоорієнтована модель (Tabular Column-oriented), документоорієнтована модель (Document-Oriented), модель, заснована на графах (Graph Databases) [11], [12], [24]. Кожна з перерахованих моделей даних характеризується структурними та технологічними особливостями, з якими варто ознайомити студентів. Проте, на відміну від реляційних моделей, реалізація однакового типу моделі бази NoSQL може мати суттєві технологічні особливості і навіть структурні відмінності в залежності від СКБД, що її підтримує. Тому вивчення конкретного типу моделі даних необхідно розглядати разом з відповідною СКБД.

Бази даних NoSQL, першочергово, призначені для реалізації різного роду Інтернет-проектів, які характеризуються надвеликими обсягами неструктурованих чи слабо структурованих даних, постійний і одночасний доступ до яких може виконувати велика кількість користувачів. Так, наприклад, кількість користувачів соціальної мережі Facebook, яка працює на основі колонкоорієнтованої моделі бази даних Apache Cassandra, становила на початок 2019 року 2,27 млрд. (за даними Digital 2019 reports from Hootsuite and We Are Social [25]).

На фоні активного розвитку баз даних NoSQL розробники реляційних баз даних вжили заходів щодо використання досвіду та наробок NoSQL у своїх продуктах, що призвело до появи так званих «нових реляційних баз даних» (NewSQL). Мотивами, що спонукали до відродження реляційних баз даних, вважаємо: відсутність у NoSQL змоги підтримки транзакцій з виконанням ACID-вимог, проблеми забезпечення цілісності даних та їх безпеки, відсутність підтримки спільної роботи з реляційними базами даних. Отже, NewSQL можна вважати третім етапом еволюції баз даних. Вони підтримують онлайн-обробку транзакцій (OLTP – Online Transaction Processing) та наближаються за рівнем продуктивності до баз даних NoSQL [26].

Сам термін «NewSQL» було запропоновано Matthew Aslett у 2011 році [27] в контексті підкреслення того факту, що відтепер існують нові покращенні реляційні бази даних, які підтримують масштабування та забезпечують високопродуктивну роботу з великими масивами даних. Робочою назвою цих нових реляційних баз даних був термін «ScalableSQL», який згодом було змінено на NewSQL як більш простий для сприйняття та розуміння.

NewSQL – це реляційні бази даних, які підтримують транзакції з виконанням ACID-вимог та використовують окремі підходи баз даних NoSQL для забезпечення горизонтального масштабування та підвищення власної продуктивності.

У таблиці 1 представлено порівняльну характеристику основних типів баз даних.

Таблиця 1

Порівняльна характеристика сучасних типів баз даних

Параметри порівняння	Реляційні БД	NoSQL	NewSQL
1	2	3	4
Дані	структуровані	неструктуровані	структуровані та неструктуровані
Масштабування	вертикальне	горизонтальне	горизонтальне
Модель представлення	таблична (реляційна)	ключ/значення, таблична колонкоорієнтована,	характерні для SQL та NoSQL

даних		документоорієнтована, графова	
Підтримка мови SQL	+	-	+
Транзакційні вимоги	ACID	BASE	ACID
Робота із Big Data	низька продуктивність	висока продуктивність	наближається до рівня NoSQL
Робота в хмарах (Cloud)	частково	підтримує	підтримує

Отже, як видно із представленого матеріалу та даних табл. 1, сучасними тенденціями розвитку баз даних можна вважати NoSQL та NewSQL. Далі наведемо аргументи щодо вибору СКБД для формування навчальної програми дисципліни «Організація баз даних NoSQL».

3. МЕТОДИКА ДОСЛІДЖЕННЯ

Вибір баз даних та систем їх керування можна виконувати за численними критеріальними показниками, основними з яких є такі: операційне середовище, у якому може працювати СКБД; тип логічної моделі бази даних; масштабованість; мова запитів; підтримка WEBтехнологій; вартість ліцензії; продуктивність та багато інших [6]. Їх використання має сенс при обґрунтуванні вибору СКБД для розробки конкретних проєктів автоматизації [11].

При виборі СКБД з точки зору організації навчального процесу потрібно керуватися дещо іншими показниками, наприклад, зрозумілість інтерфейсу та доступність мови програмування (мови запитів), відповідність обмеженим технічним параметрам комп'ютерної лабораторії, універсальність (підтримка реалізації різних предметних областей), наявність у СКБД необхідних для навчального процесу функціональних властивостей тощо.

Зрозуміло, що обрати конкретну СКБД для використання в освітньому процесі, яка б відповідала всім показникам одночасно – практично неможливо. Тому було проведено опитування (у формі анкетування) фахівців Центру інформаційно-обчислювальних та мережевих систем і представників науково-педагогічного колективу Навчально-наукового інституту «Інститут інформаційних технологій» ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана», які працюють з базами даних або викладають дисципліни, що пов'язані з ними. Метою проведеного опитування було встановлення найбільш значущих критеріальних показників, які варто брати до уваги під час добору сучасних типів баз даних для введення їх до навчальної програми дисципліни «Організація баз даних NoSQL».

Респондентам пропонувалось оцінити важливість показника за шкалою від 1 до 5 балів, де оцінці «5» відповідає найвищий рівень важливості, а оцінці «1» – найнижчий.

Отримані усереднені показники проведеного опитування представлено на рис. 1.

У результаті опрацювання даних опитування було встановлено, що найбільш важливими показниками добору баз даних та СКБД для освітніх потреб є: розповсюдженість або поширеність їх використання на практиці (середній бал, отриманий показником, становить 5 балів, тобто усі опитані визначили його як дуже важливий); інформаційна підтримка, тобто якість і повнота представленої супровідної документації (за усередненими оцінками показник отримав 4,6 бали); вартість або тип

ліцензії (середній бал показника – 4,3); відповідність програмним та апаратним вимогам, як і попередній показник, набрав 4,3 бали за результатами опитування; підтримка різних мов програмування – 3,2; підтримка вебтехнологій – 2,9. Інші показники отримали нижчі оцінки ($\leq 2,5$). Наприклад, показник продуктивності за усередненими оцінками отримав 2,5 бали, а показник простоти та зрозумілості інтерфейсу – 2,1 бали тощо.



Рис. 1. Результати проведеного опитування (усереднені показники у балах)

Отже, було виділено два найбільш вагомні критерії добору СКБД з точки зору використання в освітньому процесі:

- 1) технологійний критерій, який має такі показники:
 - розповсюдженість (поширеність);
 - відкритість ліцензії;
 - апаратна невимогливість;
- 2) організаційно-дидактичний критерій, що містить:
 - методичну та освітню підтримку (якість та повнота документації);
 - підтримку різних мов програмування;
 - підтримку вебтехнологій.

Ідентифіковані критерії та показники було покладено в основу механізму добору баз даних та СКБД для освітніх потреб.

4. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Вибір моделей та СКБД, що підтримують бази даних NoSQL і NewSQL, для введення їх до навчальної програми дисципліни «Організація баз даних NoSQL» базувався на їх відповідності двом критеріям (технологічному та організаційно-дидактичному) та шістьом показникам, які були визначені під час проведеного опитування. Розглянемо їх більш детально:

1. *Розповсюдженість*. Студенти під час навчального процесу мають отримувати актуальні знання та навички, тому під час вибору СКБД враховували ступінь їх поширеності та практичного застосування.

2. *Відкритість ліцензії*. Більшість ЗВО, через ряд об'єктивних факторів, не в змозі в повному обсязі забезпечити високий рівень матеріально-технічної бази комп'ютерних класів та дослідницьких лабораторій. Додаткові витрати на забезпечення навчального процесу пропрієтарним програмним забезпеченням є небажаними. Крім того, як зазначено в роботі [28], відкритий програмний код та можливість його зміни/модифікації є додатковим стимулом для набуття майбутніми фахівцями у сфері комп'ютерних наук фахових компетентностей.

3. *Апаратна невимогливість*. У контексті зазначеного у п. 2 апаратні вимоги до програмного забезпечення в комп'ютерних класах та лабораторіях також відіграють важливу роль у доборі СКБД, з якою будуть працювати студенти на заняттях.

4. *Підтримка* (освітня, методична та ін.). Опанування студентами основ роботи з новою базою даних забезпечується не тільки фаховими знання викладачів щодо роботи з нею, але й використанням в освітньому процесі відповідних наочних, методичних та навчальних матеріалів. Оскільки дисципліна є новою, то природно, що підручниками студенти не забезпечені. У цьому контексті важливим є вивчення представленої розробниками супровідної документації, різноманітного інформаційного та освітнього матеріалу в мережі Інтернет.

5. *Підтримка різних мов програмування*. Важливість даного показника для підготовки фахівців з комп'ютерних наук є очевидною. З організаційно-дидактичної точки зору необхідним є узгодження наявних у студентів знань мов програмування з тими мовами, які функціонально підтримує та чи інша СКБД. Наприклад, знання базових основ JAVA дадуть студентам змогу швидко зрозуміти логіку синтаксису JSON, формат якого вважається стандартом обміну даними в сучасних базах даних різного типу.

6. *Підтримка вебтехнологій*. Це сучасний тренд у розвитку ІТ, без навичок роботи з вебтехнологіями та вмінь розробки веборієнтованих інформаційних систем не можна говорити про кваліфікованість фахівця з комп'ютерних наук. Зараз горизонтальна масштабність сучасних типів баз даних досягається, окрім іншого, шляхом реалізації веборієнтованих архітектур (коли взаємодія користувача з базою даних відбувається через веббраузер), оскільки вони є більш зручними в організації багатокористувацького доступу.

Одразу відзначимо, що позитивним моментом переважної більшості NoSQL, з точки зору їх використання в освітньому процесі, є те, що вони мають відкритий код (належать до категорії відкритого програмного забезпечення), перебувають у вільному доступі і не потребують платних користувацьких ліцензій.

Для вибору СКБД був вивчений та проаналізований рейтинг баз даних, який проводить інформаційне видання DB-Engines [29]. Це видання щомісячно, починаючи з 2013 року, публікує рейтинги відомих баз даних відповідно до їх поточної популярності (розповсюдженості) у світі. У дослідженнях DB-Engines, окрім загального рейтингу СКБД, який охоплює всі типи моделей баз даних, надається рейтинг систем у межах кожного типу моделі бази даних та досліджується динаміка змін у популярності СКБД. Так, згідно з даними DB-Engines, станом на березень 2019 року на ринку ІТ було представлено 345 систем (баз даних) (рис. 2).

Лідерами традиційно є реляційні бази даних (Oracle, MySQL і Microsoft SQL Server). Нашу увагу привернув той факт, що серед NoSQL до найбільш поширених баз даних, за оцінками експертів, належать: документоорієнтована СКБД MongoDB (5

місце), СКБД Redis, що підтримує модель ключ/значення – (8 місце) та колонкоорієнтована СКБД Cassandra (11 місце).

Аналіз та вивчення СКБД MongoDB, як документоорієнтованої системи, не викликав питань щодо її введення до навчальної програми. Крім того, що MongoDB має дуже високий рейтинг серед СКБД (див. рис. 2), аналіз Інтернет-джерел показав наявність великої кількості матеріалів та онлайн курсів, які можуть бути використані студентами за відсутності рекомендованих Міністерством освіти і науки України підручників та навчальних посібників [30], [31], [32].

345 systems in ranking, March 2019

Rank			DBMS	Database Model	Score		
Mar 2019	Feb 2019	Mar 2018			Mar 2019	Feb 2019	Mar 2018
1.	1.	1.	Oracle +	Relational, Multi-model	1279.14	+15.12	-10.47
2.	2.	2.	MySQL +	Relational, Multi-model	1198.25	+30.96	-30.62
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model	1047.85	+7.79	-56.94
4.	4.	4.	PostgreSQL +	Relational, Multi-model	469.81	-3.75	+70.46
5.	5.	5.	MongoDB +	Document	401.34	+6.24	+60.82
6.	6.	6.	IBM Db2 +	Relational, Multi-model	177.20	-2.23	-9.47
7.	↑9.	7.	Microsoft Access	Relational	146.20	+2.18	+14.26
8.	↓7.	8.	Redis +	Key-value, Multi-model	146.12	-3.32	+14.90
9.	↓8.	9.	Elasticsearch +	Search engine, Multi-model	142.79	-2.46	+14.25
10.	10.	↑11.	SQLite +	Relational	124.87	-1.29	+10.06
11.	11.	↓10.	Cassandra +	Wide column	122.80	-0.58	-0.69

Рис.2. Рейтинг систем керування базами даних за версією інформаційного видання DB-Engines, березень 2019 року [29]

MongoDB – це СКБД із відкритим вихідним кодом, написана мовою програмування C++ [33]. Компанія MongoDB Inc (раніше – 10gen) вийшла з першою версією MongoDB на ринок у серпні 2009 року.

Основними елементами документоорієнтованої системи керування базами даних MongoDB є документи та їх колекція. Документ – це набір полів довільної структури, колекція – це сукупність структурно чи семантично подібних документів. Колекції не мають обмежень на вміст, а також обмежень щодо структури (різна колекція може містити різну кількість документів, різного обсягу та рівня наповненості). Ключове правило створення документу в колекції – наявність ключа (`_id`), який є унікальним у межах даної колекції. Відповідно документи ідентифікуються за допомогою назви колекції разом зі значенням `_id`. Дана СКБД має широкий арсенал інструментів роботи з документами (запити вибірки, сортування, агрегування тощо).

Робота MongoDB базується на використанні JSON (внутрішні дані зберігаються і передаються у двійковій формі – BSON), що забезпечує гнучкість моделі даних СКБД, високу продуктивність обробки даних та зрозумілість мови взаємодії для користувачів.

MongoDB «не розуміє» мову SQL, не підтримує жорстко заданих схем, у ній відсутній механізм посилкової підтримки транзакцій (починаючи з версії 4.0, MongoDB підтримує багатодокументарні транзакції, що відповідають вимогам ACID), вона добре масштабується за допомогою шардінгу (sharding) і забезпечує опрацювання Big Data на кластерах, реплікації реалізовані на основі асинхронної master-slave моделі. Ця модель передбачає, що запис даних виконується в журнал операцій (oplog) виключно на головному вузлі (master або primary), а зчитування даних може відбуватись як з головного, так і підлеглих (slave або secondary) вузлів.

Розроблено значну кількість графічних інтерфейсів, що підтримують роботу MongoDB, серед яких можемо виділити: Compass, RoboMongo, RockMongo, NoSQLBooster та інші.

MongoDB підтримує роботу в більшості операційних середовищах, зокрема з операційними системами Windows, Linux, MacOS, Solaris.

Наступною NoSQL, на яку доцільно звернути увагу, є нереляційна СКБД Redis (REmote DIctionary Server), що підтримує модель ключ/значення. Загалом база даних типу ключ/значення є досить простою в проектуванні та створенні. Основними структурними елементами є первинний ключ та значення, що асоціюється з ключем. На відміну від реляційних моделей баз даних, значення – це агрегат, що може зберігати дані довільної структури. Значення в моделі Key/Value є непрозорими, тому запити з пошуку значень певних агрегатів не можна реалізувати в цій базі даних. Можливі лише пошук значення по ключу, записування нового значення чи вилучення його з відповідним ключем. Такі бази часто називають сховищами даних. Цю назву не слід ототожнювати з поняттям класичних сховищ даних (Data Warehouse), які використовуються в аналітичних OLAP-технологіях. Термін «сховище даних», у контексті баз даних NoSQL, пов'язаний лише з функцією ефективного зберігання неструктурованих даних, без розкриття їх семантики.

Враховуючи, що в моделях типу ключ/значення завжди виконується доступ за первинним ключем, вони характеризуються високою швидкістю реалізації запитів і добре масштабуються. Це досить поширена модель даних і на сьогодні налічується понад п'ятдесят СКБД баз даних ключ/значення. Прикладами найбільш рейтингових СКБД цього типу є такі: Redis, MemcacheDB, Hazelcast, Aerospike, Ehcache, Riak KV [34].

При вивченні баз даних типу ключ/значення рекомендується знайомство з групою резидентних NoSQL in-memory баз даних (in-memory database, IMDB). Це бази даних, які характеризуються такою технологічною особливістю, як зберігання та швидкісне опрацювання даних в оперативній пам'яті. Використання таких баз даних дозволяє збільшити продуктивність опрацювання даних при зменшенні витрат на вартість самої СКБД. Швидкість опрацювання даних в оперативній пам'яті іноді в тисячі разів перевищує швидкість опрацювання на твердих носіях [35].

Враховуючи той факт, що оперативна пам'ять має фізичні обмеження, варто пам'ятати, що бази даних in-memory підходить далеко не для всіх задач. Бази типу СКБД Redis найбільш підходять для додатків, що читають і записують дані, які потрібно швидко знаходити за допомогою ключа. NoSQL in-memory бази даних використовуються найчастіше для зберігання даних:

- сесій і профілів користувачів;
- під час розробки комп'ютерних ігор;
- з управління об'єктами, де є критичним час реакції системи (телекомунікаційні системи; управління транспортними перевезеннями, різного роду безперервні технологічні процеси і т.п.);
- з організації online торгів (цінними паперами, валютою і т.п.);
- про події чи пристрої Інтернет-речей (IoT).

Для вивчення табличних колонкоорієнтованих моделей варто зосередити увагу на особливостях побудови моделі в середовищі обраної СКБД. На сьогодні найбільш поширеними СКБД даного типу є: Cassandra, HBase, Microsoft Azure Table Storage, Accumulo, Google Cloud Bigtable [36].

Колонкоорієнтована модель бази даних подібна до реляційної моделі з табличним представленням даних. Проте, на відміну від реляційних моделей, дані в яких зберігаються в рядках таблиці, що складаються з логічно пов'язаних атрибутів,

представлених різними типами й форматами, в табличних колонкоорієнтованих базах окремими рядками є дані колонок чи їх сімейств. Водночас кожна колонка містить дані одного типу і формату.

Вивчення табличних колонкоорієнтованих моделей баз даних пропонується виконувати з урахуванням особливостей СКБД Cassandra [37].

Структурними елементами моделі бази даних СКБД Cassandra є колонка (column), сімейство колонок (column family) та набір (простір) ключів (keyspace). Колонка – це основний структурний елемент, який містить: ім'я (column name), значення (value) та часову мітку (timestamp), яка може вказувати на термін актуальності даних. Значення однієї колонки представлено даними одного типу і формату. Сімейство колонок – це об'єднання декількох семантично та логічно взаємозв'язаних колонок, які читаються й адмініструються як єдине ціле. Рядок чи запис (row) – це поіменована колекція колонок, які об'єднуються за допомогою ключа (row key). Набір (простір) ключів – група ключів багатьох сімейств ключів, як правило обмежених однією предметною областю.

Слід звернути увагу на те, що колонкоорієнтована модель, як і попередні розглянуті моделі баз даних, не є універсальною. Кращим варіантом використання цього класу моделей є опрацювання великих масивів даних з метою агрегування даних та формування аналітичних звітів, а також для зберігання результатів індексації сайтів.

Наступним типом моделей баз даних, який пропонується для вивчення студентам комп'ютерних наук, є графові моделі баз даних. Вони використовуються в предметних областях, що характеризуються численними навігаційними зв'язками між даними, які слід враховувати під час їх опрацювання. Особливо поширені графові моделі баз даних у різних соціальних проєктах, зокрема соціальних мережах.

Щодо графових моделей баз даних було прийнято рішення зосередити увагу на базах даних типу NewSQL, оскільки вони є сучасними реляційними СКБД, у яких доопрацьовано функціонал і які, крім іншого, підтримують графові моделі даних.

Для вивчення графових моделей баз даних була обрана СКБД Microsoft SQL Server 2017 [38]. Вибір Microsoft SQL Server 2017 обумовлений тим, що студенти знайомляться з цією СКБД у курсі «Організація баз та сховищ даних». Отже, не потрібно витратити час на опанування технологією роботи з базою даних в середовищі даної СКБД, а можна зосередити увагу на моделюванні графів та роботі з ними. Функції графової моделі представлення даних SQL Server повністю інтегровані в ядро СКБД. Завдяки такій інтеграції є можливість користуватися широким спектром компонентів SQL Server, зокрема індексами, службами машинного навчання та ін. Для роботи з графами в SQL Server 2017 стандартними способами створюється користувацька база даних, тобто не потрібні якісь спеціальні прийоми чи налаштування БД для створення графа.

Для моделювання графів у SQL Server 2017 використовуються таблиці двох спеціальних видів. Один вид – для зберігання даних вузлів, а другий – для відображення ребер (зв'язків). Таблиці-вузли містять дані сутностей предметної області, а таблиці-ребра визначають характер та напрям зв'язку між вузлами. Таблиця-вузол графа створюється, як і реляційна таблиця, оператором CREATE TABLE, але в кінці додається параметр AS NODE. Коли вказано цей параметр, ядро бази ідентифікує таблицю як вузол. Таблиця-вузол обов'язково повинна мати поле первинного ключа.

Створення таблиць-ребер аналогічно створенню таблиць-вузлів, але замість AS NODE потрібно вказувати параметр AS EDGE. Первинний ключ в таблиці-ребрі є необов'язковим. Обов'язково потрібно вказувати ідентифікатор таблиці-вузла, з якого виходить ребро, та ідентифікатор таблиці-вузла, у який ребро входить. Інші додаткові користувацькі атрибути, що характеризують ребро є необов'язковими.

Створивши графову базу даних, студенти зможуть відпрацювати основні запити до бази даних на мові T-SQL з використанням оператора SELECT та функції MATCH, яка задає шаблон для пошуку вузлів згідно їх зв'язків і виконує навігацію по графу.

5. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Вивчення тенденцій розвитку баз даних та систем їх керування є надзвичайно важливим у сучасних умовах, коли стрімкий розвиток нових технологій призводить до різкого збільшення обсягів даних, зміни структури даних та ступеня їх однорідності. Критичним моментом опрацювання великих масивів даних є бази даних. Ці дані не можуть зберігатися та опрацьовуватися з використанням традиційних реляційних баз даних, які зорієнтовані на підтримку структурованої інформації. Забезпечувати ефективне зберігання та розподілене швидкісне опрацювання великих обсягів неструктурованих даних стало можливим завдяки появі нових типів баз даних, як-от NoSQL та NewSQL. Саме вивчення актуальних типів баз даних та систем їх керування в підготовці студентів комп'ютерної спеціальності є необхідною та обґрунтованою відповіддю на сучасні тенденції в розвитку інформаційних технологій. Ознайомлення студентів з базами даними NoSQL та NewSQL, принципами роботи з актуальними (з точки зору ринку та роботодавця) СКБД призведе до підвищення якості підготовки майбутніх фахівців з комп'ютерних наук.

З'ясовано, що при введення СКБД до навчальної програми дисципліни доцільно звертати увагу на її рейтинг (поширеність використання), якість і повноту супровідної документації (рівень інформаційної та методичної підтримки), операційне середовище, апаратні вимоги, необхідність придбання ліцензії. У результаті аналізу сучасних баз даних та систем їх керування на предмет відповідності зазначеним критеріям до навчальної програми дисципліни «Організація баз даних NoSQL» було додано бази даних NoSQL (для вивчення моделі даних ключ/значення – СКБД Redis, для вивчення документоорієнтованої моделі даних – СКБД MongoDB, для вивчення колонкоорієнтованої моделі даних – СКБД Cassandra) та NewSQL (для вивчення графової моделі даних – СКБД Microsoft SQL Server 2017).

Бази даних NoSQL і NewSQL є прогресивними напрямками розвитку ІТ, актуальними для провідних країн світу. Проте вже зараз існують нові концепції та підходи до проектування та функціонування баз даних нового типу. Зокрема широке поле для досліджень відкриває нова концепція мультисховищ (polystore databases) як перспективний напрям розвитку баз даних. Перспективними є подальші дослідження, що пов'язані з вивченням та аналізом функціоналу, ліцензійної політики нових систем і можливості їх введення до навчальних програм для подальшого поглибленого вивчення. Одним із важливих аспектів, що потребує подальших наукових досліджень, є добір сучасних типів баз даних згідно з ідентифікованими в роботі критеріями та показниками за методом експертного оцінювання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Кабінет Міністрів України. (2018, Січ. 17). Розпорядження № 67-р, «Про схвалення Концепції розвитку цифрової економіки та суспільства України на 2018-2020 роки та затвердження плану заходів щодо її реалізації». [Електронний ресурс]. Доступно: <http://zakon5.rada.gov.ua/laws/show/67-2018-%D1%80>. Дата звернення: Лютий 21, 2019.
- [2] Сайт кафедри системотехніки Харківського національного університету радіоелектроніки. [Електронний ресурс]. Доступно: http://sedep.nure.ua/?page_id=147. Дата звернення: Лютий 24, 2019.

- [3] Сайт Національного технічного університету «Київський політехнічний інститут ім. Ігоря Сікорського» спеціальність 122 «Комп'ютерні науки». [Електронний ресурс]. Доступно: <http://arpers.kpi.ua/122>. Дата звернення: Лютий 24, 2019.
- [4] Сайт кафедри інформаційних систем в економіці Інституту інформаційних технологій в економіці Київського національного економічного університету ім. Вадима Гетьмана. [Електронний ресурс]. Доступно: https://fisit.kneu.edu.ua/ua/depts9/k_informacijnyh_system_v_ekonomici. Дата звернення: Березень 13, 2019.
- [5] К. Дж. Дейт, *Введение в системы баз данных*. Пер. с англ. Киев, Украина: Диалектика, 1998.
- [6] Н.В. Ситник, *Проективання баз і сховищ даних*. Київ, Україна: КНЕУ, 2004.
- [7] М. Фаулер, Дж. Садаладж Прамодкумар, *NoSQL: новая методология разработки нереляционных баз данных*. Пер. с англ. Москва, Россия: ООО «И.Д. Вильямс», 2013.
- [8] Я. Робинсон, Д. Вебер, Э. Эфрем, *Графовые базы данных: новые возможности для работы со связанными данными*. Пер. с англ. Москва, Россия: ДМК Пресс, 2016.
- [9] І.Б. Швороб, «Методи та засоби екстракції та аналізу слабоструктурованих текстових даних на основі документо-орієнтовного графа», дис. канд. наук., Нац. ун-т «Львівська політехніка», Львів, 2018.
- [10] Paolo Atzenin, Francesca Bugiotti, Luca Rossi, «Uniform access to NoSQL systems», *Information Systems*, p. 1-17, 2013. doi: <http://dx.doi.org/10.1016/j.is.2013.05.002>. Accessed on: March 04, 2019.
- [11] João Ricardo Lourenzol, Bruno Cabral, Paulo Carreiro, Marco Vieira, and Jorge Bernardino, «Choosing the right NoSQL database for the job: a quality attribute evaluation», *Journal of Big Data*, p. 1-26, 2015. [Електронний ресурс]. Доступно: <https://link.springer.com/article/10.1186/s40537-015-0025-0>. Дата звернення: Берез. 14, 2019.
- [12] Ahmad Abdullah, Qingfeng Zhuge, «From Relational Databases to NoSQL Databases: Performance Evaluation», *Research Journal of Applied Sciences, Engineering and Technology*, no. 11(4), p. 434-439, 2015.
- [13] В.О. Брацький, О.М. М'якшило, «Дослідження особливостей застосування реляційних і нереляційних баз даних на прикладі SQL Server та MongoDB», *Наукові праці Національного університету харчових технологій*, Т. 22, № 5, с. 15-24, 2016.
- [14] С.О. Спасітелева, В.Л. Бурячок, «Комплексний захист гетерогенних корпоративних сховищ даних», *Сучасний захист інформації*, №1, с. 58-65, 2017.
- [15] Vatika Sharma, Meenu Dave, «SQL and NoSQL Databases», *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 8, p.20-28, 2012.
- [16] М.Ю. Швець, Д.С. Заруба, Ю.В. Хохлов, «Порівняння SQL та NoSQL баз даних», *Вчені записки ТНУ імені В.І. Вернадського*. Серія: технічні науки, Т. 29 (68), Ч. 2, № 6, с. 21-25, 2018.
- [17] Sneha Binani, Ajinkya Gutti, Shivam Upadhyay, «SQL vs. NoSQL vs. NewSQL- A Comparative Study», *Communications on Applied Electronics (CAE) – Foundation of Computer Science FCS*, New York, USA, vol. 6, no.1, p. 43-46, 2016.
- [18] Jeremy Kepner et al., «Associative Array Model of SQL, NoSQL, and NewSQL Databases», *2016 IEEE High Performance Extreme Computing Conference (HPEC)*, Waltham, MA, USA, 2016. [Електронний ресурс]. Доступно: <https://arxiv.org/ftp/arxiv/papers/1606/1606.05797.pdf>. Дата звернення: Берез. 10, 2019.
- [19] List of NoSQL Databases. [Електронний ресурс]. Доступно: <http://nosql-database.org/>. Дата звернення: Лют. 24, 2019.
- [20] Jeyhun Karimov, Tilmann Rabl, and Volker Markl, «PolyBench: The First Benchmark for Polystores», in *10th TPC Technology Conference «Performance Evaluation and Benchmarking for the Era of Artificial Intelligence»*, Rio de Janeiro, 2018, p. 24-41 [Електронний ресурс]. Доступно: http://www.redaktion.tu-berlin.de/fileadmin/fg131/dima-feed/polystore_benchmark_TPCTC-1028_cv.pdf. Дата звернення: Берез. 01, 2019.
- [21] Т. Коннолли, К. Бегг, А. Страчан. *Базы данных: проектирование, реализация и сопровождение. Теория и практика*. Пер. с англ. Москва, Россия: Издательский дом «Вильямс», 2000.
- [22] Дж. Мартин, *Организация баз данных в вычислительных системах*. Пер. с англ. Москва, Россия: Мир, 1980.
- [23] Keith D. Foote, «A Brief History of Non-Relational Databases», 2018. [Електронний ресурс]. Доступно: <https://www.dataversity.net/a-brief-history-of-non-relational-databases/>. Дата звернення: Лют. 21, 2019.
- [24] Н.В. Ситник, «Роль і місце баз даних в цифровій економіці», на *I наук.-метод. конф. Цифрова економіка*, Київ, 2018, с. 267-258.
- [25] Digital 2019: Global internet use accelerates. [Електронний ресурс]. Доступно: <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>. Дата звернення: Лют. 27, 2019.

- [26] Ahmed Almassabi, Omar Bawazeer, and Salahadin Adam, «Top NewSQL databases and features classification», *International Journal of Database Management Systems*, Vol.10, no.2, p. 11-31, 2018.
- [27] Matthew Aslett, «What we talk about when we talk about NewSQL», 2011. [Електронний ресурс]. Доступно: https://blogs.the451group.com/information_management/2011/04/06/what-we-talk-about-when-we-talk-about-newsql. Дата звернення: Лют. 25, 2019.
- [28] І.С. Зінов'єва, В.О. Артемчук, А.В. Яцишин, «Використання відкритих геоінформаційних систем у підготовці фахівців з комп'ютерних наук», *Інформаційні технології і засоби навчання*, Том 68, №6, с. 87-99, 2018. [Електронний ресурс]. Доступно: <https://journal.iitta.gov.ua/index.php/itlt/article/download/2567/1415>. Дата звернення: Лютий 20, 2019.
- [29] DB-Engines Ranking, 2019. [Електронний ресурс]. Доступно: <https://db-engines.com/en/ranking>. Дата звернення: March 14, 2019.
- [30] Начало работы с MongoDB, часть 1. [Електронний ресурс]. Доступно: <http://ashep.org/2012/nachaloraboty-s-mongodb-chast-1/#.XILDv8kzZkh>. Дата звернення: Лют. 24, 2019.
- [31] Начало работы с MongoDB, часть 2. [Електронний ресурс]. Доступно: <http://ashep.org/2012/nachaloraboty-s-mongodb-chast-2/#.XILDv8kzZkh>. Дата звернення: Лют. 24, 2019.
- [32] Онлайн-руководство по MongoDB. [Електронний ресурс]. Доступно: <https://metanit.com/nosql/mongodb>. Дата звернення: Лют. 27, 2019.
- [33] MongoDB: Open Source Document Database, 2019. [Електронний ресурс]. Доступно: <https://www.mongodb.com>. Дата звернення: Берез. 01, 2019.
- [34] DB-Engines Ranking of Key-value Stores, 2019. [Електронний ресурс]. Доступно: <https://db-engines.com/en/ranking/key-value+store>. Дата звернення: Берез. 01, 2019.
- [35] Как выбрать In-memory NoSQL базу данных с умом. Тестируем производительность. [Електронний ресурс]. Доступно: <https://habr.com/ru/company/oleg-bunin/blog/312062>. Дата звернення: Берез. 06, 2019.
- [36] DB-Engines Ranking of Wide Column Stores. [Електронний ресурс]. Доступно: <https://db-engines.com/en/ranking/wide+column+store>. Дата звернення: Берез. 9, 2019.
- [37] Apache Cassandra, 2019. [Електронний ресурс]. Доступно: <http://cassandra.apache.org>. Дата звернення: Берез. 9, 2019.
- [38] Создание базы данных графа и выполнение некоторых шаблоны запросов с помощью T-SQL [Електронний ресурс]. Доступно: <https://docs.microsoft.com/ru-ru/sql/relational-databases/graphs/sql-graph-sample?view=sql-server-2017>. Дата звернення: Берез. 18, 2019.

Матеріал надійшов до редакції 11.04.2019 р.

СОВРЕМЕННЫЕ БАЗЫ ДАННЫХ NoSQL В ПОДГОТОВКЕ БАКАЛАВРОВ СПЕЦИАЛЬНОСТИ "КОМПЬЮТЕРНЫЕ НАУКИ"

Ситник Нина Васильевна

кандидат экономических наук, доцент, профессор кафедры информационных систем в экономике
ГБУЗ «Киевский национальный экономический университет имени Вадима Гетьмана», г. Киев, Украина
ORCID ID 0000-0002-4032-4881
vfsytnyk@gmail.com

Зиновьева Ирина Сергеевна

кандидат экономических наук, доцент кафедры информационных систем в экономике
ГБУЗ «Киевский национальный экономический университет имени Вадима Гетьмана», г. Киев, Украина
ORCID ID 0000-0001-5122-8994
ira.zinovyeva@kneu.edu.ua

Аннотация. Статья посвящена проблеме овладения современными подходами к организации баз данных будущими специалистами по специальности 122 «Компьютерные науки». Базы данных NoSQL и NewSQL – новейшие тренды в сфере IT, они требуют детального ознакомления и изучения студентами компьютерных специальностей. В статье исследованы и раскрыты понятия баз данных NoSQL и NewSQL. Проведена сравнительная характеристика реляционных баз данных с базами данных типа NoSQL, NewSQL. В связи с широким распространением различного вида WEB-сервисов, потребностью распределенной обработки больших массивов данных (Big Data) обоснована необходимость полноценного изучения нового поколения баз данных и получения профессиональных компетентностей по их практическому применению. Изучение новейших баз данных не может быть

альтернативой изучению реляционных баз данных. Учитывая концептуальные различия и технологические особенности использования современных баз данных, в Киевском национальном экономическом университете имени Вадима Гетьмана в учебные планы специальности 122 «Компьютерные науки» включена новая дисциплина «Организация баз данных NoSQL», изучение которой пополнит знания студентов об основных тенденциях развития баз данных и повысит их конкурентоспособность на рынке труда. При разработке учебной программы дисциплины были проанализированы учебные планы подготовки специалистов компьютерных наук ведущих заведений высшего образования Украины, литературные и интернет-источники, которые подтвердили необходимость и целесообразность изучения баз данных NoSQL и NewSQL. Исследование критериев и анализ рейтинга систем управления базами данных позволил сделать вывод о целесообразности включения в рабочую программу дисциплины «Организация баз данных NoSQL» ряда наиболее распространенных в мировой практике баз данных типа NoSQL (MongoDB, Redis, Cassandra) и NewSQL (Microsoft SQL Server 2017), которая имеет расширенный функционал и поддерживает работу с графами).

Ключевые слова: реляционные базы данных; NoSQL и NewSQL; системы управления базами данных; компьютерные науки.

MODERN NoSQL DATABASES FOR TRAINING BACHELORS OF “COMPUTER SCIENCE” SPECIALTY

Nina V. Sytnyk

PhD in Economics, Associate Professor, Professor at Department of Information Systems in the Economics
Kyiv National Economic University named after Vadym Hetman, Kyiv, Ukraine
ORCID ID 0000-0002-4032-4881
vfstytnyk@gmail.com

Iryna S. Zinovieva

PhD in Economics, Associate Professor at Department of Information Systems in the Economics
Kyiv National Economic University named after Vadym Hetman, Kyiv, Ukraine
ORCID ID 0000-0001-5122-8994
ira.zinovyeva@kneu.edu.ua

Abstract. The article deals with students' training in modern approaches to the organization of databases. NoSQL and NewSQL databases are the latest trends in IT field, they should be learnt and mastered by computer sciences specialists during their study. In the article the notions of databases NoSQL and NewSQL are explored and defined. The authors compare relational databases with NoSQL and NewSQL. Due to the wide distribution of different types of WEB-services, as well as the need for distributed big data processing, the article substantiates the necessity for a thorough study of a new generation of databases and professional competencies concerning their practical application acquiring. However, the authors state that studying the latest databases cannot be an alternative to studying the relational databases. Considering the conceptual differences and technological peculiarities of using the newest databases in Kyiv National Economic University named after Vadym Hetman a new course on organization of NoSQL databases has been introduced in the curriculum of specialty 122 «Computer Sciences». The course aims at deepening students' knowledge about current main trends in the databases development and developing their competitiveness in the labor market. During the formulation of the discipline curriculum, there have been analyzed the educational plans of computer sciences specialists training of leading Ukrainian institutions of higher education, literary and Internet sources, which confirmed the necessity and appropriacy for studying the NoSQL and NewSQL databases. Criteria research and analysis of the database management systems rating made it possible to make a conclusion about the practicability of including in the curriculum of *Organization of NoSQL Databases* course a number of the most widespread databases of the NoSQL (MongoDB, Redis, Cassandra) and NewSQL (Microsoft SQL Server 2017 DBMS), which has advanced functionality and supports work with graphs.

Keywords: relational databases; NoSQL and NewSQL; database management systems; computer sciences.

REFERENCES (TRANSLATED AND TRANSLITERATED)

- [1] Cabinet of Ministers of Ukraine. (2018, Jan. 17). Order № 67-r, «About the approval of the Concept for the development of the digital economy and society in Ukraine for 2018-2020 and approval of the plan of measures for its implementation». [Online]. Available: <http://zakon5.rada.gov.ua/laws/show/67-2018-%D1%80>. Accessed on: February 21, 2019 (in Ukrainian).
- [2] Site of the Department of system engineering by Kharkiv National University of Radio Electronics [Online]. Available: http://sedep.nure.ua/?page_id=147. Accessed on: February 24, 2019 (in Ukrainian).
- [3] The site of the National Technical University of Ukraine «Kyiv Polytechnic Institute. Igor Sikorsky» for specialty 122 «Computer Science». [Online]. Available: <http://apeps.kpi.ua/122>. Accessed on: February 24, 2019 (in Ukrainian).
- [4] Site of the Department of information systems in the economist of the Institute of Information Technologies in the economy of the Kyiv National Economic University named after Vadym Hetman. [Online]. Available: https://fisit.kneu.edu.ua/ua/depts9/k_informacijnyh_system_v_ekonomici. Accessed on: March 13, 2019 (in Ukrainian).
- [5] K. Dzh. Deit, *Introduction to database systems*. Translate from English. Kyiv, Ukraine: Dyalektyka, 1998 (in Russian).
- [6] N.V. Sytnyk, *Designing databases and data warehouses*. Kyiv, Ukraine: KNEU, 2004 (in Ukrainian).
- [7] M. Fowler, Dzh. Sadaladzh Pramodkumar, *NoSQL: New Non-Relational Database Development Methodology*. Translate from English. Moscow, Russia: LTD «Y.D. Vyliams», 2013 (in Russian).
- [8] Ya. Robynson, D. Veber, E. Eyfrem, *Graph databases: new features for working with related data*. Translate from English. Moscow, Russia: DMK Press, 2016 (in Russian).
- [9] I.B. Shvorob, «Methods and means of extraction and analysis of poorly structured text data on the basis of a document-oriented graph», dys. kand. nauk., Nats. un-t «Lvivska politekhnika», Lviv, 2018 (in Ukrainian).
- [10] Paolo Atzenin, Francesca Bugiotti, Luca Rossi, «Uniform access to NoSQL systems», *Information Systems*, pp. 1-17, 2013.doi: <http://dx.doi.org/10.1016/j.is.2013.05.002>. (in English).
- [11] João Ricardo Lourenzol, Bruno Cabral, Paulo Carreiro, Marco Vieira, and Jorge Bernardino, «Choosing the right NoSQL database for the job: a quality attribute evaluation», *Journal of Big Data*, pp. 1-26, 2015. [Online]. Available: <https://link.springer.com/article/10.1186/s40537-015-0025-0>. Accessed on: March 14, 2019 (in English).
- [12] Ahmad Abdullah, Qingfeng Zhuge, «From Relational Databases to NoSQL Databases: Performance Evaluation», *Research Journal of Applied Sciences, Engineering and Technology*, no. 11(4), pp. 434-439, 2015 (in English).
- [13] V.O. Bratskyi, and O.M. Miakshylo, «Investigation of the use of relational and non-relational databases on the example of SQL Server and MongoDB», *Naukovi pratsi Natsionalnoho universytetu kharkovoykh tekhnolohii*, v. 22, no 5, pp. 15-24, 2016 (in Ukrainian).
- [14] S.O. Spasitieliava, and V.L. Buriachok, «Integrated protection of heterogeneous corporate data warehouse», *Suchasnyi zakhyst informatsii*, no.1, pp. 58-65, 2017 (in Ukrainian).
- [15] Vatika Sharma, Meenu Dave, «SQL and NoSQL Databases», *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 8, pp.20-28, 2012 (in English).
- [16] M.Iu. Shvets, D.S. Zaruba, and Yu.V. Khokhlov, «Compare SQL and NoSQL databases», *Vcheni zapysky TNU imeni V.I. Vernadskoho*. Series: technical science, vol. 29 (68), ch. 2, no 6, pp. 21-25, 2018 (in Ukrainian).
- [17] Sneha Binani, Ajinkya Gutti, and Shivam Upadhyay, «SQL vs. NoSQL vs. NewSQL- A Comparative Study», *Communications on Applied Electronics (CAE) – Foundation of Computer Science FCS*, New York, USA, vol. 6, no.1, pp. 43-46, 2016 (in English).
- [18] Jeremy Kepner et al., «Associative Array Model of SQL, NoSQL, and NewSQL Databases», *2016 IEEE High Performance Extreme Computing Conference (HPEC)*, Waltham, MA, USA, 2016. [Online]. Available: <https://arxiv.org/ftp/arxiv/papers/1606/1606.05797.pdf>. Accessed on: March 10, 2019 (in English).
- [19] List of NoSQL Databases. [Online]. Available: <http://nosql-database.org/>. Accessed on: February 24, 2019 (in English).
- [20] Jeyhun Karimov, Tilmann Rabl, and Volker Markl, «PolyBench: The First Benchmark for Polystores», in *10th TPC Technology Conference «Performance Evaluation and Benchmarking for the Era of Artificial Intelligence»*, Rio de Janeiro, 2018, pp. 24-41 [Online]. Available: http://www.redaktion.tu-berlin.de/fileadmin/fg131/dima-feed/polystore_benchmark_TPCTC-1028_crv.pdf. Accessed on: March 01, 2019 (in English).
- [21] T. Konnelly, K. Behh, and A. Strachan, *Databases: design, implementation and maintenance. Theory and practice*. Translate from English. Moscow, Russia: Yzdatelskyi dom «Vyliams», 2000 (in Russian).

- [22] Dzh. Martyn, *Database organization in computing systems*. Translate from English. Moscow, Russia: Myr, 1980 (in Russian).
- [23] Keith D. Foote, «A Brief History of Non-Relational Databases», 2018. [Online]. Available: <https://www.dataversity.net/a-brief-history-of-non-relational-databases/>. Accessed on: February 21, 2019 (in English).
- [24] N.V. Sytnyk, «The role and place of databases in the digital economy», on *I nauk.-metod. conf. Digital economy*, Kyiv, 2018, pp. 267-258 (in Ukrainian).
- [25] Digital 2019: Global internet use accelerates. [Online]. Available: <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>. Accessed on: February 27, 2019 (in English).
- [26] Ahmed Almassabi, Omar Bawazeer, and Salahadin Adam, «Top NewSQL databases and features classification», *International Journal of Database Management Systems*, vol.10, no.2, pp. 11-31, 2018 (in English).
- [27] Matthew Aslett, «What we talk about when we talk about NewSQL», 2011. [Online]. Available: https://blogs.the451group.com/information_management/2011/04/06/what-we-talk-about-when-we-talk-about-newsql. Accessed on: February 25, 2019 (in English).
- [28] I.S. Zinovieva, V.O. Artemchuk, and A.V. Iatsyshyn, «The use of open geoinformation systems in computer science education», *Information Technologies and Learning Tools*, vol. 68, no.6, pp. 87-99, 2018. [Online]. Available: <https://journal.iitta.gov.ua/index.php/itlt/article/download/2567/1415>. Accessed on: February 25, 2019 (in Ukrainian)
- [29] DB-Engines Ranking, 2019. [Online]. Available: <https://db-engines.com/en/ranking>. Accessed on: March 14, 2019 (in English).
- [30] Getting Started with MongoDB, Part 1. [Online]. Available: <http://ashep.org/2012/nachalo-raboty-s-mongodb-chast-1/#.XILDv8kzZkh>. Accessed on: February 24, 2019 (in Russian).
- [31] Getting Started with MongoDB, Part 2. [Online]. Available: <http://ashep.org/2012/nachalo-raboty-s-mongodb-chast-2/#.XILDv8kzZkh>. Accessed on: February 24, 2019 (in Russian).
- [32] Online MongoDB Guide. [Online]. Available: <https://metanit.com/nosql/mongodb>. Accessed on: February 27, 2019 (in Russian).
- [33] MongoDB: Open Source Document Database, 2019. [Online]. Available: <https://www.mongodb.com>. Accessed on: March 01, 2019 (in English).
- [34] DB-Engines Ranking of Key-value Stores, 2019. [Online]. Available: <https://db-engines.com/en/ranking/key-value+store>. Accessed on: March 01, 2019 (in English).
- [35] How to choose an in-memory NoSQL database wisely. Testing performance. [Online]. Available: <https://habr.com/ru/company/oleg-bunin/blog/312062>. Accessed on: March 06, 2019 (in Russian).
- [36] DB-Engines Ranking of Wide Column Stores. [Online]. Available: <https://db-engines.com/en/ranking/wide+column+store>. Accessed on: March 9, 2019 (in English).
- [37] Apache Cassandra, 2019. [Online]. Available: <http://cassandra.apache.org>. Accessed on: March 9, 2019 (in English).
- [38] Creating a graph database and executing some query templates using T-SQL [Online]. Available: <https://docs.microsoft.com/ru-ru/sql/relational-databases/graphs/sql-graph-sample?view=sql-server-2017>. Accessed on: March 18, 2019 (in Russian).

