

СУЧАСНІ ПІДХОДИ ДО ПОДАЛЬШОЇ ЕВОЛЮЦІЇ КОНЦЕПЦІЇ БАЗ ДАНИХ

Зінов'єва Ірина Сергіївна

к.е.н., доцент, доцент кафедри ІСЕ

ДВНЗ «Київський національний економічний університет імені Вадима
Гетьмана»

Артемчук Володимир Олександрович,

к.т.н, старший науковий співробітник

старший науковий співробітник відділу
моделювання енергетичних процесів і систем

Інституту проблем моделювання в
енергетиці ім. Г.Є. Пухова НАН України

Анотація. У статті представлено спробу систематизації та узагальнення існуючих типів та видів баз даних, розглянуто основні підходи до класифікації баз даних та наведено загальну їх характеристику. Визначення основних типів та видів баз даних наведено у хронологічному порядку їх розвитку.

Ключові слова: бази даних, класифікаційні ознаки, моделі представлення даних, реляційні, нереляційні, багатомодельні бази даних, мультисховища

Активний розвиток сучасних інформаційних технологій та мережі Інтернет призвели до ускладнення вимог до організації та функціонування баз даних, як одного із базових інструментів функціонування інформаційних систем та програмних продуктів.

На сучасному ринку баз даних представлено значну кількість видів та типів баз даних, розроблено значну кількість класифікаційних ознак [1] (див. табл. 1). Наприклад, поширеним є підхід до класифікації баз даних за їх вмістом, в межах якого прийнято поділяти бази даних типи, в залежності, від сфери їх

використання: геоінформаційні бази даних (GIS database), виступають основою функціонування геоінформаційних систем [2, 3].

Таблиця 1

Класифікація сучасних баз даних

№ з/п	Класифікаційна ознака	Види
1	За кількістю користувачів:	• бази даних, що підтримують роботу одночасно одного користувача; • бази даних, що підтримують роботу одночасно багатьох користувачів;
2	За технологією обробки:	• централізовані; • децентралізовані (розподілені);
3	За способом доступу до даних:	• локальні; • мережеві;
4	За типом моделі представлення даних:	• реляційні; • об'єктно-орієнтовні; • нереляційні; • мережеві; • багатомодельні; • мультисховища
5	За технологією фізичного зберігання:	• в оперативній пам'яті; • у вторинній пам'яті (на жорстких дисках); • на переносних зчитуючих пристроях;
6	За вмістом:	• геоінформаційні; • наукові; • мультимедійні;

		• мобільні; • часові; • інтернет речей (IoT)
7	За типом використовуваного програмного забезпечення для управління даними:	• гомогенні; • гетерогенні;
8	За типом ліцензії:	• відкриті; • умовно-відкриті; • комерційні;
9	За технологією підтримки передачі та оброблення даних:	• OLTP; • OLAP; • робота з великими даними та аналітика; • XML; • JSON.

У цих базах даних міститься різного роду просторова інформація, в т.ч. координати геопросторового розміщення об'єктів на земній поверхні, описи атрибутів, векторні і растрові дані та інше; наукові бази даних (databases for science research), використовуються під час та для проведення наукових досліджень, пошуку наукової інформації, статей, метаданих тощо; мультимедійні бази даних (multimedia database) містять різного типу контент: текстовий, аудіо, відео та анімацію, графічні об'єкти тощо; мобільні бази даних (mobile databases) використовуються для роботи на мобільних пристроях (планшети, смартфони), згідно чого можуть легко переноситись з одного пристрою на інший, забезпечують обробку локальних запитів, підтримую роботу в режимі реального часу (real-time); бази даних часових рядів (time series database) – призначені для роботи із даними часової динаміки (хронологією); бази даних інтернету речей (Database for IoT & Sensor Data) призначені для

керування даними, отриманими від сенсорів, датчиків, з пристроїв зчитування біометричних та телематичних даних.

Розглянута класифікаційна ознака, як інші, на думку авторів, є не ключовою у виділенні типів та видів сучасних баз даних. В якості найбільш унікальної класифікаційної ознаки варто обрати типи моделей представлення даних, оскільки за іншими класифікаційними ознаками бази даних можуть бути реалізовані на основі підтримки тієї чи іншої моделі даних. З іншого боку, ознака моделі представлення даних дає змогу чітко систематизувати усі наявні види та типи баз даних (див. рис. 1).



Рис. 1. Таксономія класифікації сучасних баз даних

Починаючи з 50-х років і до кінця XX ст. концепції організації баз даних еволюціонували від файлових систем збереження даних на основі використання перфокарт до об'єктно-орієнтованих та традиційних реляційних баз даних. Найбільш стабільні позиції та високий рівень затребуваності серед користувачів і розробників на кінець XX ст. отримали традиційні бази даних, які через підтримку мови запитів SQL та реляційної моделі представлення даних отримали назву баз даних SQL. Це цілий клас баз даних та систем керування базами даних, які активно використовуються і наразі (Microsoft SQL, Oracle, MySQL, Postgres, IBM Db2 тощо).

Більшість наявних на ринку баз даних SQL відноситься до типу пропрієтарного (закритого, захищеного комерційною ліцензією) програмного забезпечення, тобто для його використання необхідно придбати у вендора ліцензію на

використання бази даних. Серед найпоширеніших можна визначити Oracle, Microsoft SQL Server, IBM Db2, Microsoft Access. Частка відкритих (open source databases) баз даних SQL є значно меншою, ніж пропрієтарних на ринку, проте останнім часом їх кількість має тенденцію до зростання. Найбільш поширеними вважаються MySQL, SQLite та PostgreSQL [4].

Розвиток інтернету, поява мови розмітки XML у 90-х роках XX ст. та масовий перехід користувачів на роботу з веб-технологіями призвели до накопичення в інформаційних системах та сховищах даних напівструктурованих та неструктурованих даних, що значно ускладнювало роботу реляційних баз даних. У цей час з'явився новий вид баз даних – бази даних XML (NXD), які підтримували структуру та мову запитів XML – як базову. Прикладами цього виду баз даних можна вважати MarkLogic, Virtuoso, Sedna, searchxml, eXist-db. Але, як і об'єктно-орієнтовані бази даних, цей вид не набув широкого поширення і замінити собою бази даних SQL. Тим не менш, істотна частка сучасних реляційних баз даних використовують та підтримують мову XML. Зокрема, у п'ятій версії стандарту SQL: 2003 окремий розділ присвячено специфікації SQL / XML, який визначає правила взаємодії між схемами XML і структурами SQL та функції обробки XML-даних через SQL-запити [5].

На поч. XIX ст. відбулось переосмислення концепції реляційних баз даних на фоні появи нових методів добування даних, розвитку концепцій сховищ даних, а також через проблеми баз даних SQL щодо роботи із неструктурованими даними, складністю агрегування, невисокою продуктивністю роботи та ін. Альтернативною концепцією баз даних стала концепція нереляційних або NoSQL баз даних.

На противагу традиційним реляційним базам даних, бази даних NoSQL здатні забезпечувати високу продуктивність, горизонтальну масштабованість та гнучкість в роботі з неструктурованими та значними за обсягами даними. На даному етапі, концепцію баз даних NoSQL підтримують такі типи баз даних: сховища ключ/значення (key-value stores), колонко-орієнтовані сховища (wide column stores), документарні сховища (document stores), графові бази даних

(graph databases) [6]. Як підвид графових баз даних виділяють бази даних RDF, які використовують для зберігання метаданих та представлення їх у вигляді семантичної павутини. До числа типових представників цього підвиду графових бази даних відносять: Apache Jena-TDB, Redland, RDF4J, CubicWeb.

Деякі науковці схильні виділяти ще один тип – пошукові бази даних (search engines database), хоча їх відношення до типів NoSQL є умовним. З одного боку, пошукові бази даних не мають жорстких структурних вимог до збереження та представлення даних, що характерно для NoSQL, з іншого – у них використовується індексування та мова запитів, що сході на SQL.

Майже усі бази даних NoSQL мають відкриту ліцензію та відносяться до open source databases. Окрім того, бази даних NoSQL прийнято поділяти за типом та середовищем обробки даних на: бази даних, що операції виконують в оперативній пам'яті системи (in-memory), бази даних, що працюють із використанням SSD або жорсткими дисками (HDD). Варто додати, що значна частка баз даних NoSQL активно використовуються у своїй роботі хмарні веб-технологіями, працюють в режимі real-time та підтримують обмін даними через XML та JSON. До найбільш поширених баз даних NoSQL можна віднести: MongoDB, Riak, Redis, Splunk, Cassandra, Hbase, Neo4j, Elasticsearch. Актуальними проблемами сучасних баз даних NoSQL вважаються проблеми узгодженості даних, функціональної обмеженості в реалізації складних запитів, точність результатів обробки.

Поява концепції NoSQL та затребуваність нереляційних баз даних на ринку призвела до того, що розробники реляційних баз даних почали використовувати елементи концепції та підходи по роботі з великими неструктурованими даними у свої базах даних, які отримали назву NewSQL або нові реляційні бази даних. Вони краще масштабуються, ніж традиційні реляційні бази, мають вищу продуктивність роботи, забезпечують високу якість в роботі з даними за рахунок підтримки онлайн-транзакцій. Бази даних NewSQL підтримують обмін різними видами даних, роботу із хмарними веб-технологіями, працюють у режимах real-time та streaming, мають здатність

фіксувати потокові та просторові дані, підтримують обмін даними через XML, JSON та ін. Прикладами баз даних типу NewSQL можна вважати: ClustrixDB, VoltDB, SQL Server, NuoDB, CockroachDB.

На разі на ринку програмного забезпечення та інформаційних систем існує значна конкуренція між реляційною, NoSQL та NewSQL концепціями баз даних. В результаті цієї конкуренції почали з'являтися бази даних, в яких реалізується підтримка як реляційних так і нереляційних моделей даних [7]. При чому, таке поєднання відбувається не на рівні технологій роботи із даними, хоча і це має місце, а на рівні розробки так званого «мультисховища» – надбудови, яка може керувати одночасно базами даних з реляційними та нереляційними моделями даних. Таким чином, сучасними підходами до подальшої еволюції концепції баз даних є:

1. Багатомодельні бази даних (multi-model databases) реалізуються на базі використання одного єдиного інтегрованого сервера, який забезпечує підтримку кількох типів моделей даних. У табл. 2 наведено загальний перелік сучасних багатомодельних баз даних із зазначенням типів даних, які вони підтримують.

Таблиця 2

Характеристика підтримки моделей представлення даних у сучасних багатомодельних базах даних

Назва	Тип	Підтримувана модель даних									
		реляційна	ключ/ значення	колонко- орієнтована	документарна	графова	RDF	пошукова	NXD	часові ряди	об'єктно- орієнтована
Oracle	SQL				+	+	+				
MySQL	SQL				+						
Microsoft SQL Server	SQL				+	+					
IBM Db2	SQL				+		+				
MongoDB	NoSQL							+			

	(документарна)										
Redis	NoSQL (ключ/значення)				+	+		+		+	
Kdb+	NoSQL (пошук ова)	+									
Amazon DynamoDB	NoSQL (ключ/значення)				+						
ArangoDB	NoSQL		+		+	+		+			
Sqrrl	NoSQL		+	+	+	+					
MariaDB	NewSQL				+	+					
Teradata	NewSQL				+	+				+	
SAP HANA	NewSQL				+	+					
Oracle Berkeley DB	багато модель на		+						+		
Ignite	багато модель на	+	+								
Microsoft Azure	багато		+	+	+	+					

Cosmos DB	модель на										
InterSystems Cache	багато модель на	+	+		+						+
Virtuoso	багато модель на	+			+	+	+		+		
Amazon Neptune	багато модель на					+	+				
InterSystems IRIS	багато модель на	+	+		+						+
MarkLogic	багато модель на				+		+	+	+		

Наприклад, Oracle Database – система управління базами даних, що базується на багатомодельному підході, підтримує роботу з реляційними, графовими та колонко-орієнтованими моделями даних, ArangoDB – забезпечує підтримку моделей ключ/значення, документо-орієнтованої та графової моделі даних. Як видно із представлених даних, можна виділити два основні види багатомодельних баз даних: 1) бази даних побудовані на основі підтримки конкурентної моделі даних SQL (або NoSQL, або NewSQL) і доповнені окремими типами підтримки інших моделей даних; 2) бази даних побудовані на основі підтримки різних моделей даних. До цього виду відносяться такі бази даних, які мають рівноцінний функціонал, що характерний для як для NewSQL, так і для NoSQL. Загалом, багатомодельність значно спрощує інтеграцію даних в середині сховища;

2. Мультисховища (polystore databases) – являють собою так звану «вітрину даних», яка забезпечує доступ інтегрованих між собою баз даних на основі прямого доступу та використанні локальних мов запитів (характерних для конкретної моделі даних).

Концепція організації мультисховища полягає у використанні різних типів баз даних для обробки різних потреб на основі проміжного програмного забезпечення – інтерпретаторів. З одного боку, використання мультисховищ звільняє користувача розуміти специфіку чи особливості роботи із тим чи іншим типом баз даних, знати усі специфічні мови запитів, способи обробки даних тощо. З іншого боку, така архітектура створює нові проблеми організації доступу до даних та їх інтеграції. Наразі, концепція мультисховищ реалізована у таких продуктах: Mугia, BigDAWG, Polypheny-DB.

Отже, на початковому етапі розвитку концепцій організації баз даних серед ієрархічної, мережевої та реляційної концепції, найбільш поширеною стала остання, завдяки чому майже протягом сорока років на ринку не було альтернатив реляційним базам даних. В результаті суцільної цифровізації суспільства можливостей реляційних баз даних виявилось недостатньо, що стало поштовхом для подальшої еволюції концепцій. У XXI ст. задоволення вимог швидкості оброблення інформації, гнучкості систем управління, адаптивності моделей даних, масштабованості забезпечують із різним рівнем ефективності концепції NoSQL, NewSQL, багатомодельних баз даних та мультисховищ.

СПИСОК ЛІТЕРАТУРИ

1. Types And Classification Of Database Management System [Online]. Available: <https://whatisdbms.com/wp-content/uploads/2017/03/Types-of-DBMS.pdf>. Accessed on: September 24, 2019.
2. Яцишин А.В., Попов О.О., Артемчук В.О., Ковач В.О. та Зінов'єва І.С., «Автоматизовані інформаційні системи підтримки прийняття управлінських

рішень у галузі екологічної безпеки», *Information Technologies and Learning Tools*, Vol 72, No 4, С. 286–305, 2019.

3. Zinovieva I., Artemchuk V., Iatsyshyn A., «The use of open geoinformation systems in computer science education», *Information Technologies and Learning Tools*, No 68(6), pp. 87-99, 2018.

4. DB-Engines Ranking, 2019. [Online]. Available: <https://db-engines.com/en/ranking>. Accessed on: October 31, 2019

5. George Feuerlicht, «Database Trends and Directions: Current Challenges and Opportunities», *Databases, Texts, Specifications, Objects*, vol. 567, pp. 163–174, 2010.

6. Kristi L. Berg, Tom Seymour, Richa Goel, «History Of Databases», *International Journal of Management & Information System*, Vol.17, p. 29-35, 2013.

7. 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.