

БАЗИ ДАНИХ NOSQL I NEWSQL В СИСТЕМАХ BIG DATA

Н.В. Ситник, к.е.н., доцент

Київський національний економічний університет імені Вадима Гетьмана

Останнє десятиліття характеризується вибуховим зростанням обсягів даних і появою нового терміну великі дані (BigData). Вузким місцем в опрацюванні та аналізі великих даних є бази даних. Більше тридцяти років в сфері ІТ-технологій домінували реляційні бази даних. Проте на сьогоднішній день реляційні моделі та системи керування базами даних (СКБД), що їх підтримують, не відповідають цілому ряду вимог, основними з яких є наступні:

- Потреба підтримки та опрацювання агрегованих, неструктурованих даних. Реляційні моделі підтримують лише структуровані атомарні дані.
- Горизонтальне масштабування та робота з кластерами. Реляційні СКБД масштабуються вертикально шляхом збільшення потужностей серверів. Таке масштабування є вартісним і має суто технічну межу.
- Підтримка постійних змін даних та забезпечення одночасного доступу великої кількості користувачів. Для реляційних баз даних, що підтримують транзакції шляхом виконання ACID-вимог (Atomicity - атомарність, Consistency - узгодженість, Isolation - ізолюваність, Durability – довговічність зберігання), характерним є певні часові затримки при внесенні змін до бази даних.

Для оброблення великих даних застосовуються бази даних NoSQL та NewSQL. Термін NoSQL розшифровується як NotOnlySQL — «не лише SQL», що не заперечує реляційні бази даних. Порівняльні характеристики реляційних та баз даних NoSQL наведені в таблиці 1. Жодна зі спільних характеристик не реляційних баз даних, що наведені в таблиці 1, не є визначальною, що свідчить про нечіткість визначення терміну NoSQL[2].

Не дивлячись на досить широке поширення, бази даних NoSQL мають суттєвий недолік, який є перешкодою для їх використання в корпоративних та бізнесових системах. Цей недолік полягає в тому, що NoSQL системи не

підтримують транзакції, а задовольняють BASE-вимоги (Basically Available - базова доступність, Soft state— гнучкий стан, Eventually consistent – узгодженість в підсумку), що негативно впливає на точність результатів опрацювання даних.

Таблиця 1.

Реляційні бази даних	Бази даних NoSQL
Теоретичний базис – теорія нормалізації та реляційна алгебра	Теорема CAP
Атомарні дані (aggregate-ignore)	Агреговані дані довільної структури
Структуровані нормалізовані дані з гарантованою їх консистентністю	Неструктуровані денормалізовані слабо консистентні дані
Підтримка діаграми (схеми) даних	Відсутність статичної схеми
Мова запитів SQL	Відсутність мови SQL
Вертикальне масштабування	Горизонтальне масштабування
Підтримка транзакцій та ACID-вимог	Підтримка BASE-вимог
Інтегроване рішення для традиційних OLTP-задач: бухгалтерський облік, фінанси, виробництво, банківська справа і т.п.	Кожна модель бази NoSQL орієнтована під конкретний тип задач з опрацювання BigData. Таких як: WED-сервіси, мобільні додатки, мультимедіа, геоінформаційні системи, системи штучного інтелекту і т.п.

Вирішенням проблеми слабкої узгодженості даних є бази даних NewSQL[3]. NewSQL—це реляційні бази даних, що підтримують розподілену архітектуру та роботу з кластерами, мають механізм підтримки транзакцій і реплікацій та мову SQL. Відомі компанії розробники реляційних СКБД

Microsoft, Oracle і DB2 доопрацювали свій функціонал, який дозволяє підтримувати бази NewSQL.

В ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана» до навчального плану освітньо-професійної програми спеціальності 122 «Комп'ютерні науки» включено нову дисципліну «Організація баз даних NoSQL». Розробка програми даної дисципліни вимагала вирішення цілого ряду питань. Перш за все виникало питання, які моделі баз даних та СКБД потребують першочергового вивчення. На сьогодні налічується більше ніж 225 СКБД, що підтримують бази даних NoSQL[4]. Аналіз рейтингу баз даних [1] дозволив дійти висновку, що найбільш поширеними моделями бази даних NoSQL є такі: ключ/значення (Key/Value), таблична колонко-орієнтована (Tabular Column-oriented), документо-орієнтована (Document-Oriented), модель основана на графах (Graph Databases). Для вивчення документо-орієнтованих, табличних колонко-орієнтованих та моделі типу ключ/значення вибрано відповідні найбільш рейтингові СКБД: MongoDB, Cassandra та Redis. Для баз даних NewSQL вибрана СКБД Microsoft SQL Server 2017, яка підтримує модель основу на графах.

Вивчення баз даних NoSQL є необхідною складовою підготовки спеціалістів з комп'ютерних наук, що сприятиме розширенню їх фахових компетенцій та підвищенню конкурентоспроможності на ринку праці. Крім того опанування баз даних NoSQL створить теоретичну основу для дисципліни «Технології оброблення та аналіз надвеликих масивів даних».

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. DB-Engines Ranking. [Електронний ресурс]. Доступно: <https://db-engines.com/en/ranking>. Дата звернення: 11.03. 2019.

2. Фаулер, Мартин, Садаладж, Прамодкумар Дж. NoSQL: новая методология разработки реляционных баз данных. : Пер. с англ. - М.: ООО "И.Д. Вильямс", 2013. - 192 с.
3. NewSQL — новый виток в эволюции BigData, забираем лучшее из SQL и NoSQL.[Електронний ресурс]. Доступно:<https://habr.com/ru/post/154589/>. Дата звернення: 09.03. 2019.
4. List of NoSQL Databases. [Електронний ресурс]. Доступно: <http://nosql-database.org/>. Дата звернення: 24.02. 2019.

СТАН ВИКОРИСТАННЯ BIGDATA У НАВЧАЛЬНОМУ ПРОЦЕСІ ЗВО УКРАЇНИ

В.В. Кулаженко, к.е.н., ст. викладач

Київський національний торговельно-економічний університет

На стан сучасної університетської освіти впливає декілька факторів, основними з яких являються глобалізаційні процеси, виникнення нових знань і компетентностей, демографічні проблеми, впровадження новітніх інформаційних технологій у побутове життя, а також можливості збору та аналізу BigData(Великих даних).

На жаль, заклади вищої освіти в Україні недостатньо використовуються Великі дані у навчальному процесі. Причинами цього є не підготовленість викладачів до використання Великих даних, застаріла матеріально-технічна база, недостатньо швидкі темпи впровадження Великих даних у роботу бізнесових та державних структур.

За визначенням компанії Gartner, першого розробника BigData, Великі дані – це інформаційні активи великого обсягу, високої швидкості, та/або високої різноманітності, які вимагають нових форм обробки, щоб дозволити здійснювати покращення прийняття рішень, відкриття ідей і оптимізацію процесів.