

УДК 004.62:004.65:004.75

*Алексюк В. В., здобувач 4 курсу спеціальності 122 Комп'ютерні науки,
Комаров В. Ф., канд. техн. наук, старший викладач кафедри інформаційних
технологій*

MONGODB ЯК ЕФЕКТИВНИЙ ЗАСІБ У ЗАДАЧАХ ОБРОБКИ ТА АНАЛІЗУ ДАНИХ

Донецький національний університет імені Василя Стуса, м. Вінниця

У міру того, як компанії намагаються подолати все більший обсяг і різноманітність інформаційних потоків, потреба у гнучких, масштабованих і швидких системах баз даних (БД) стає все більш очевидною. У динамічному середовищі науки про дані (Data Science), де попит на ефективні рішення для управління даними та аналітики продовжує невпинно зростати, MongoDB, що пропонує універсальну платформу та знаходиться на старті вирішення означених проблем, стає ключовим гравцем, який змінює загальний підхід до великих наборів даних (Big Data) та їх використання.

Існує два основні типи баз даних: реляційні (SQL) та нереляційні (NoSQL). Реляційні бази даних зберігають дані у стовпцях і рядках. Microsoft SQL Server, Oracle і Sybase використовують реляційні системи керування базами даних (СКБД). На відміну від них, бази NoSQL зберігають неструктуровані дані без схем у численних колекціях і вузлах. Вони не потребують фіксованих схем таблиць, масштабуються горизонтально, але підтримують обмежену кількість запитів на об'єднання. MongoDB є представником таких БД.

MongoDB була створена в 2009 р. [1] як відкрита, масштабована, надійна і безкоштовна база даних NoSQL з відкритим вихідним кодом, зарекомендувала себе як універсальна, гнучка база даних, і сьогодні використовується як внутрішнє сховище даних багатьох відомих компаній та організацій, як-от Forbes, Facebook, Google, IBM, X (Twitter) та інші.

Актуальність MongoDB в контексті Data Science складно переоцінити. Її пристосованість до роботи з неструктурованими та напівструктурованими даними в поєднанні з можливістю безперешкодної інтеграції з популярними мовами програмування позиціонує її як потужний інструмент в арсеналі фахівців.

MongoDB має такі ключові особливості [2]:

- **загальне призначення** – MongoDB може обслуговувати різноманітні набори даних та різні цілі в межах однієї бази даних;
- **гнучка модель** – MongoDB використовує гнучку, безсхемну модель документів, що дає змогу зберігати різні типи даних в одній колекції;
- **масштабованість** – MongoDB розроблена для горизонтального масштабування, що робить її добре придатною для задач із швидко зростаючими обся-

гами даних, розподіляючи дані між декількома серверами; MongoDB може забезпечувати безперебійну масштабованість для пристосування до зростаючих навантажень;

➤ **мова запитів** – MongoDB допомагає користувачам виконувати складні запити, агрегації та індексування завдяки потужній та виразній мові запитів, що полегшує ефективний пошук та аналіз даних і має вирішальне значення для задач науки про дані;

➤ **JSON** – для бази даних використовується той самий протокол, що широко використовується для зв'язку між інтерфейсом і API.

Завдяки гнучкій моделі документів, що підтримує різні типи даних, масштабованості та можливостям мови запитів MongoDB є динамічним та універсальним рішенням для використання в задачах Data Science, яке дає такі можливості [3]:

➤ **Зберігання та обробка даних.** MongoDB слугує основним сховищем для зберігання різноманітних даних, зокрема, неструктурованих або напівструктурованих даних, які часто зустрічаються в Data Science.

➤ **Робота з великими обсягами даних.** MongoDB масштабується горизонтально, легко обробляючи великі обсяги даних. Така масштабованість має вирішальне значення в сучасних задачах обробки даних, обсяги яких постійно збільшуються.

➤ **Аналітика та агрегація даних.** Використовуючи мову запитів та фреймворк агрегації MongoDB, складні аналітичні операції та агрегацію даних можна виконувати безпосередньо в базі даних, що суттєво спрощує процес підготовки даних до аналізу.

➤ **Геопросторовий аналіз.** Завдяки вбудованій підтримці геопросторових запитів та індексації MongoDB є ефективним інструментом для аналізу й обробки географічних даних.

➤ **Зберігання та використання результатів моделей машинного навчання.** MongoDB може слугувати сховищем для вхідних даних та результатів моделей машинного навчання, даючи змогу зберігати й керувати моделями в єдиному середовищі.

➤ **Інтеграція з інструментами Data Science.** MongoDB легко інтегрується з різними популярними інструментами Data Science та мовами програмування, як-от Python або R, забезпечуючи безперебійний робочий процес та обмін даними між середовищами.

Останній пункт варто розібрати детальніше. Python, що є популярною мовою програмування у спільноті фахівців Data Science, легко інтегрується з MongoDB за допомогою бібліотеки PyMongo [4]. Бібліотека надає інтерфейс Python для взаємодії з MongoDB, який дає змогу швидко отримати доступ для

виконання маніпуляцій із даними, їх пошуку і аналізу, коли на встановлення з'єднання з базою даних потрібно лише кілька рядків коду.

Такий зв'язок Python та MongoDB стає все популярнішим, оскільки обидва інструменти є вкрай потужними в області розробки та обробки даних. Python як мова програмування володіє зручним синтаксисом, що робить його обраною мовою для великої кількості розробників та аналітиків. MongoDB як БД типу NoSQL пропонує гнучку модель документів, що ідеально взаємодіє з динамічним та розширюваним характером Python. Це допомагає розробникам та аналітикам з легкістю працювати з неструктурованими даними, а також швидко й ефективно взаємодіяти з базою даних. Швидкість і гнучкість є ключовими факторами поєднання Python та MongoDB.

Підсумовуючи вкажемо, що інтеграція MongoDB у Data Science виходить за межі традиційного зберігання даних, слугуючи рушійною силою для інновацій. Окрім своєї фундаментальної ролі в управлінні даними, MongoDB надає переваги у вирішенні безпосередньо задач обробки та аналізу даних, поєднання великих наборів даних і машинного навчання.

Список використаних джерел

1. Why Use MongoDB: What It Is and What Are the Benefits. URL: <https://www.simplilearn.com/tutorials/mongodb-tutorial/what-is-mongodb>
2. Growing Significance of MongoDB in Data Science Field. URL: <https://www.edureka.co/blog/the-growing-significance-of-mongodb-in-data-science-field/>
3. Introduction to MongoDB for Data Science. URL: <https://www.knowledgehut.com/blog/data-science/mongodb-for-data-science>
4. Exploring MongoDB for Data Science and Python. URL: <https://www.linkedin.com/pulse/exploring-mongodb-data-science-python-aritra-pain-3rd-year-/>

УДК 004.04

*Підруцький Д. А., здобувач 3 курсу спеціальності 122 Комп'ютерні науки,
Січко Т. В., канд. техн. наук, доцент, доцент кафедри інформаційних технологій*

BIG DATA В ОБРОБЦІ І АНАЛІЗІ ДАНИХ З РІЗНИХ СФЕР

Донецький національний університет імені Василя Стуса, м. Вінниця

У сучасному світі, охопленому хвилею технологічного розвитку, важко уявити галузь, яка залишається поза впливом великих обсягів даних, або так