

УДК 004.9

Лебідко Д.М., Онуферко В.А., Перетятко Т.П.

Тернопільський національний технічний університет імені Івана Пулюя, Україна

ВЕЛИКІ ЗА ОБСЯГОМ ДАНІ, РЕЛЯЦІЙНІ ТА НЕРЕЛЯЦІЙНІ МОДЕЛІ

Lebidko D.M., Onuferko V.A., Peretiatko T.P.

BIG DATA, RELATIONAL AND NON-RELATIONAL MODELS

В кінці минулого століття наука та технології вступили в нову еру – еру великих за обсягом даних. Науковий прогрес призвів до різкого збільшення кількості даних, що продукуються щодня. Наукові дані, зібрані на даний час, здебільшого залежать від ряду різних змінних. Тому їх можна представити у вигляді багатовимірних масивів або гіперкубів даних [1]. Впорядкування даних усередині гіперкубів є важливою задачею, і для цього було розроблено багатий спектр різноманітних інструментів.

Більшість сучасних програмних архітектур забезпечують підтримку обробки структур даних на базі гіперкубів [2]. Однак у кожній з цих програм гіперкуби даних можуть отримувати доступ до даних лише «ортогонально» до своїх «осей». Це означає, що можна вибирати лише певні значення або діапазони вздовж заданих розмірів. Це обмежує можливості аналізу даних, оскільки часто потрібно отримувати доступ до даних у неортогональних напрямках. Наприклад, може бути потрібно порівняти дані з різних осей або визначити, як зміна однієї змінної впливає на інші.

Поширення великих за обсягом даних створило потребу в масштабованих системах, які можуть ефективно їх обробляти. Реляційні бази даних, які базуються на SQL, є стандартним підходом до зберігання та керування даними. Вони добре працюють для структурованих даних, зокрема, таблиць, але мають обмеження щодо масштабованості. Бази даних NoSQL, з іншого боку, не використовують реляційну модель даних. Вони зазвичай краще підходять для керування великими обсягами даних, оскільки можуть масштабуватися по горизонталі, тобто формувати додаткові вузли для обробки даних.

Бази даних NoSQL – це нереляційні бази даних, які використовують різні моделі зберігання даних, що краще підходять для сучасних типів даних. На даний час існує чотири основні типи баз даних NoSQL: Бази даних типу «ключ-значення» зберігають дані у вигляді пари ключ-значення. Ключ ідентифікує запис, а значення містить дані про запис. Ці бази даних добре підходять для зберігання великих обсягів даних, які не мають чіткої структури. Бази даних документів зберігають дані у вигляді документів, які є JSON-подібними об'єктами. Ці бази даних добре підходять для зберігання неструктурованих даних, таких як текст і код. Бази даних стовпців зберігають дані у вигляді стовпців. Ці бази даних добре підходять для зберігання даних, які часто запитуються за певними стовпцями. Графові бази даних зберігають дані у вигляді графів. Граф складається з вузлів та ребер, які подають об'єкти та відношення між ними. Ці бази даних добре підходять для зберігання даних, що мають багато зв'язків між сутностями.

Захист особистої інформації користувачів є важливою задачею при передачі хмарних сховищ. Це пов'язано з тим, що хмарне сховище містить чутливі дані, зокрема, імена, адреси електронної пошти, номери кредитних карток та інші особисті дані.

Література

1. J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh, "Data cube: A relational aggregation operator generalizing group-by, cross-tab, and subtotals," *Data mining and knowledge discovery*, vol. 1, pp. 29–53, 1997.

2. Xtensor Stack, "Xtensor Documentation," In *Xtensor* (Version 0.24.6). Retrieved from Read the Docs: [<https://xtensor.readthedocs.io/en/latest/>], 2023.