

УДК 004.65

## ДОСЛІДЖЕННЯ МЕТОДІВ ПОРЦІЮВАННЯ БАЗ ДАНИХ ДЛЯ МАСШТАБУВАННЯ ПРОГРАМНИХ СИСТЕМ

Ковалевич Б. І.

Науковий керівник – к.т.н., доцент Мазурова О.О.

Харківський національний університет радіоелектроніки  
61166, Харків, просп. Науки, 14, каф. Програмної інженерії  
м. Харків, України

тел.: 0970233845, e-mail: bohdan.kovalevych@nure.ua

The given work is devoted to research of database partitioning methods as a method of horizontal scaling of the systems. The main target of the research is comparing methods of data partitioning for MS SQL Server and CosmosDB. In this work, it was proposed the different data partitioning strategies available in MS SQL and CosmosDB. Research queries were developed and metrics were selected to evaluate their performance. During the experiments, query execution performance was investigated for the proposed slicing schemes for MS SQL and CosmosDB with different workload characteristics.

Однією з ключових проблем створення корпоративних розподілених систем в сучасному світі це їх масштабованість. Ефективним рішенням цієї проблеми є горизонтальне масштабування таких систем та баз даних (БД). Порціювання є одним із популярних методів горизонтального масштабування БД. Отже, актуальним та важливим для практики є використання саме тих методів порціювання та відповідних систем управління базами даних (СУБД), які дозволять забезпечити необхідний рівень продуктивності програмних додатків, що працюють з БД.

Метою даної роботи є дослідити методи порціювання БД та порівняти їх продуктивність при використанні класичних реляційних БД та NoSQL БД, що останні роки активно використовуються у проектах з обробки великих об'ємів даних.

Було проведено аналіз підходів [1, 2], що використовують SQL та NoSQL СУБД до порціювання даних, та обрано СУБД для подальшого дослідження. MS SQL Server використовує метод горизонтального масштабування, який називається partition scheme. Цей метод ефективно працює, якщо дані розподілені між різними серверами у різних регіонах. Кожен сервер має свою порцію бази даних, та ці копії синхронізуються для забезпечення консистентності даних. Цей метод може бути корисним для високошвидкісних додатків, що потребують швидкого доступу до даних. CosmosDB, з іншого боку, використовує метод порціювання за сетом налаштованих ключів. Цей метод також передбачає розміщення порцій (сегментів) даних на різних серверах, які незалежні один від одного, та обробка даних в межах одного сегменту даних є більш швидкою та ефективною.

Для дослідження продуктивності методів порціювання для обраних СУБД було обрано наступні метрики: рівномірність розподілення, відповідність операцій даних для сегментів, інтеграція в існуючу структуру даних, час обчислювальних затрат, час очікування.

Для дослідження було спроектовано базу даних в сфері керування процесом працевлаштування студентів на базі бюджету держзамовлень, де відбувається велике навантаження з боку обробки даних під час обрахунку фінансових операцій та формуванням звітності. Розроблена БД містить інформацію про такі сутності, як `subsidized_placement` (місце субсидії), `delivery_program` (програма субсидії), `enrollment` (зарахування працівника), `fiscal_year` (фінансовий проміжок), деталі користувача (`user_details`), тощо.

Було розроблено схеми порціювання для обраних СУБД на основі розробленої схеми БД, а саме: для MS SQL Server метод порціювання потребував задання функції `fiscal_year_partition_func` та схеми порціювання `fiscal_year_partition_scheme`; для CosmosDB було виділено запропоновано декілька схем порціювання з використанням таких ключів, як `fiscalYearId`, `deliveryProgramId`.

У порівнянні цих методів важливо виділити можливість задати більш гнучку схему порціювання даних для MS SQL Server, а для CosmosDB важливою є можливість задати більше 1 ключа для порціювання. Дослідження продуктивності проводилося шляхом виконання серії запитів до БД, що включають в себе запити на отримання комплексних статистик на отримання звітної документації працевлаштування за обраним строком фінансовим кварталом, запити на редагування даних про місця субсидії в конкретному фінансовому кварталі, запити на видалення зарахування працівника в рамках програми субсидії.

Було проведено серію експериментів та отримані відповідні заміри для обраних метрик, що дозволило зробити висновки та сформулювати рекомендацію, щодо використання досліджених методів порціювання. Так, використовувати MS SQL Server та відповідний метод порціювання варто, якщо швидкість обробки запитів та доступу до даних є основними вимогами проекту. Порціювання та використання CosmosDB рекомендується при умові, якщо глобальний доступ до даних та низькі затримки мережі є ключовими вимогами проекту.

Список використаних джерел:

1. Daniel, A. (2009, 11 квітня). Data Partitioning. SpringerLink. [https://link.springer.com/referenceworkentry/10.1007/978-0-387-39940-9\\_688/](https://link.springer.com/referenceworkentry/10.1007/978-0-387-39940-9_688/)
2. Horizontal, vertical, and functional data partitioning - Testprep Training. Testprep Tutorials. <https://www.testpreptraining.com/tutorial/horizontal-vertical-and-functional-data-partitioning/>.