

прогнозування даних. Метод може повністю використовувати часову послідовність даних ціни акції для отримання більш надійного прогнозування.

Порівнявши критерії оцінки CLSTM з багатошаровим перцептроном (MLP), CNN, RNN, LSTM і CNN-RNN, доведено, що CLSTM має високу точність прогнозування і більше підходить для прогнозування курсу акцій.

Список літератури

1. J. H. Wang and J. Y. Leu, "Stock market trend prediction using ARIMA-based neural networks," in Proceedings of the IEEE International Conference on Neural Networks (ICNN), pp. 2160–2165, Washington, DC, USA, June 1999
2. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: <https://doi.org/10.1162/neco.1997.9.8.1735L>
3. Ho, W.-C. Kao, Y.-Y. Ou, and K.-L. Hua, "Using deep learning neural networks to predict stock market," 2019. [Online]. Available: <https://arxiv.org/abs/1903.12258>.
4. Krollner, B., Vanstone, B.J., Finnie, G.R.: Financial time series forecasting with machine learning techniques: a survey. In: ESANN, pp. 25–30 2010.
5. Bao, W., Yue, J., Rao, Y.: A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS One 12(7), e0180944 2017.
6. Baba, N., Kozaki, M.: An intelligent forecasting system of stock price using neural networks. Int. Jt. Conf. Neural Netw. 1, 371–377 1992.
7. S. M. Madge, "Predicting stock price direction using support vectormachines," 2015.
8. Zhao, Z., Zhang, X., Zhou, H., Li, C., Gong, M., Wang, Y.: Hetnerec: Heterogeneous network embedding based recommendation. Knowl. Based Syst. 204, 106218 2020.
9. I. Virtanen and P. Yli-Olli, "Forecasting stock market prices in a thin security market," Omega, vol. 15, no. 2, pp. 145–155, 1987.
10. Armano, G., M. Marchesi, and A. Murru. "A hybrid genetic-neural architecture for stock indexes forecasting." Information Sciences, 2005

УДК 004.051

*АЛЕКСАНДРОВ В. С.,
СТЕЦЕНКО І. В.*

АНАЛІТИЧНА ОБРОБКА ДАНИХ РИНКУ ОРЕНДИ НЕРУХОМОСТІ З ВИКОРИСТАННЯМ OLAP-ТЕХНОЛОГІЇ

У даному науковому дослідженні здійснено огляд специфіки роботи з великими обсягами даних/Запропоновано використання OLAP-технології, що представляє інтуїтивно зрозумілу модель даних, організовуючи їх у вигляді багатомірних кубів. Зібрано перелік рекомендацій які слід мати на увазі під час налаштування та проектування OLAP-кубів, задля забезпечення високої продуктивності і функціональності. Представлено структуру OLAP-кубу для отримання аналітичних даних про ринок оренди нерухомості.

OLTP, OLAP ТЕХНОЛОГІЇ

This research examines the specifics of working with large amounts of data. The use of OLAP technology, which provides an intuitive data model, organizing them in the form of multidimensional cubes, is proposed. Developed a list of recommendations that should be followed when setting up and designing OLAP cubes to ensure high performance. The structure of the OLAP cube for obtaining analytical data on the real estate rental market is presented.

OLTP, OLAP TECHNOLOGY

1. Вступ

Прогрес у сфері обробки та аналізу даних призводить до постійно зростаючої потреби у розробці нових технології для швидкого

доступу до значної кількості аналітичних даних. Це спричинило популярність досліджень у сфері сховищ даних, та, зокрема технології використання кубів у багатьох

сферах бізнесу. Технологія онлайн-аналітичної обробки знайшла застосування в сферах мерчандайзингу, системах прийняття рішень, сховища даних знань тощо. Бази даних, які зазвичай використовуються для зберігання всіх своїх транзакцій і записів, називаються базами даних онлайн-обробки транзакцій (OLTP). У цих базах даних зазвичай є записи, які вводяться один за одним і містять велику кількість інформації, яку можуть використовуватись для прийняття обґрунтованих рішень. Проте бази даних, які використовуються для зберігання даних, не пристосовані для цілей аналізу, адже не були розроблені окремо для такого застосунку. Тому отримання відповідей із цих баз даних є дорогим з точки зору часу та зусиль [1]. OLAP - це клас додатків та технологій, призначених для оперативної аналітичної обробки багатовимірних даних (збір, зберігання, аналіз) для аналізу діяльності корпорації та прогнозування майбутнього стану з метою підтримки прийняття управлінських рішень. Технологія OLAP застосовується, щоб спростити роботу з багатоцільовими накопиченими статистичними даними великого обсягу, а також перетворити набір кількісних показників на якісні, використовуючи швидкий, одноманітний, оперативний доступ до різноманітних форм подання інформації [2].

2. OLAP-куб

За допомогою OLAP можна аналізувати дані вздовж ієрархії вимірів на основі історичних даних. Таким чином велику кількість детальних даних можна описати більш стисло й узагальнено, що зручно для користувачів. У системі прийняття рішень дані зазвичай організовуються в куби даних. Основна функція куба даних полягає в агрегації атрибутів кількох вимірів даних. За допомогою попереднього розрахунку куба даних час запиту даних можна значно скоротити. Дані в кубі OLAP зазвичай зберігаються у схемі зірка або сніжинка. З однією або кількома централізованими таблицями фактів, кожна з яких пов'язана з однією або кількома таблицями вимірів. Таблиці фактів складаються з мір куба даних і зовнішніх ключів, що вказують на таблиці вимірів. Міри будуються на основі сховища даних. Міри в таблиці фактів називаються

групою мір. Міри – це якісна характеристика значення. Міри в таблиці фактів або у вимірі групуються в групу мір. Структура куба визначається вимірами куба даних. За допомогою вимірів, міри класифікуються таким чином, щоб їх можна було аналізувати. Різні виміри куба даних дають можливість аналізувати дані з різних точок зору. Виміри мають атрибути, які є дискретними ідентифікаторами і з яких складається вимір [3]. Елементи виміру зазвичай впорядковуються в ієрархії, щоб забезпечити структуру виміру. Часто використовуваним параметром є дата, яка може бути організована в ієрархії календаря, де день групується в місяць, місяць – у квартал, а квартал – у рік. Це дає можливість аналізувати показники на основі різних рівнів у вимірах. Вимір може мати більше однієї ієрархії, параметр дати може, наприклад, мати ієрархію календарних дат і ієрархію фінансових дат.

3. Операції над OLAP-кубами

При аналізі з використанням OLAP-кубів, основними та найбільш використовуваними операціями є згортка, розгортка та переріз. Вони використовуються для аналізу даних на різних рівнях або просто для аналізу частини куба. Операція згортки дає змогу отримати огляд даних на основі певної ієрархії вимірів. Ми можемо виконувати операцію згортки відносно параметру дати, отримуючи більш загальні результати аналізу, наприклад відобразивши дані не за окремий місяць, а на протязі року. Розгортка є протилежною до операції згортки і показує більш детальні дані. Зріз — це підмножина куба, де одне значення відповідає одному або вимірам. Наприклад, якщо з виміру часу вибрано один окремий рік, усі інші роки будуть опущені, а підкуб, що залишився, називається фрагментом вихідного куба. Операція перерізу також може бути виконана в двох або більше вимірах [3].

4. Методика створення ефективного OLAP-кубу

При розробці куба OLAP існує кілька важливих аспектів, які забезпечують функціональний і високопродуктивний куб. Перше, що слід згадати, це використання підтримуваних джерел даних і

постачальників даних. У добре спроектованому кубі OLAP дизайн вимірів є, мабуть, найважливішою частиною, і дуже важливо переконатися, що атрибути, відносини та ієрархії відповідають потребам кінцевого користувача та правильно представляють основні дані. Факти або виміри в таблиці фактів називаються групою мір. Якщо в кубі є кілька таблиць фактів, але вони зазвичай не використовуються разом, може бути гарною ідеєю створити більше одного куба, кожен з яких стосується певної теми. Це тому, що кілька груп вимірювань можуть мати негативний вплив на продуктивність запиту, навіть якщо запитується лише одна група вимірювань [4]. Якщо запит, який використовується рідше, вимагає більше одного куба, можна створити спеціальний куб шляхом створення пов'язаних груп показників. Якщо більше ніж одна група заходів базується на одному вимірі та має однакову деталізацію, вони, можливо, мають бути в одній групі мір, що містить кілька розділів. Дизайн вимірів є, мабуть, найважливішим аспектом хорошого дизайну для куба OLAP. Почнемо з того, що вимір має бути незалежним і пов'язаним з бізнес суб'єктами [5]. Важливою частиною при розробці вимірів є відносини атрибутів, усі відносини атрибутів у даних мають бути визначені в кубі, щоб дані відображали коректно. Це допомагає оптимізувати зберігання даних і обмеження MDX. Не потрібно використовувати непотрібні атрибути, оскільки вони негативно

впливають на продуктивність і зручність використання для кінцевого користувача. Часто краще додати атрибути, які визнаються необхідними, ніж спочатку додати всі знайдені атрибути та видалити непотрібні. Слід уникати зайвих зв'язків, оскільки вони лише збільшують складність і не приносять переваги вимірам. Отже, такі зв'язки повинні бути усунені, щоб уникнути проблем і підвищити продуктивність [4]. Ієрархії є важливою частиною вимірів, і щоразу, коли є атрибути, пов'язані з виміром, слід враховувати ієрархію. Усі виміри повинні мати принаймні одну ієрархію, винятком є вимір «батьківсько-дочірній», оскільки вони певним чином мають особливий тип ієрархії. Порядок атрибутів в ієрархії важливий, якщо атрибути нижнього рівня містять менше членів, ніж вище, порядок атрибутів часто знаходиться в неправильному порядку [6]. В ієрархіях всі атрибути повинні мати зв'язок з вищенаведеним рівнем, щоб мати можливість оптимізувати сервер. Куби OLAP агрегуються з даних, які зберігаються в якомусь джерелі даних, найчастіше в схемі зірка або сніжинка в сховищі даних. Зазвичай дані в сховищі даних збираються з одного або кількох джерел даних, і джерела даних можуть, швидше за все, мати не однакову структуру. бази даних. Вибрати схему зірки чи сніжинки, залежить від розміру, структури та інших характеристик даних. Специфіку використання кожної схеми представлено у таблиці 1.

Табл. 1. Порівняння схеми бази даних для формування OLAP кубу

	Схема сніжинки	Схема зірки
Обслуговування	Не має надлишкових даних, легша для обслуговування	Має надлишкові дані
Простота використання	Потребує використання складніших запитів, що робить її складнішою для розуміння та використання	Прості запити, зрозумілі для розуміння та використання
Ефективність запитів	Має більшу кількість зовнішніх ключів, отже займає більше часу при виконанні запитів	Менша кількість зовнішніх ключів, менше часу на виконання запиту
Сховище даних	Добре підходить у якості сховища даних з відношеннями багато до багатьох	Добре підходить для сховищ даних з відношеннями один до одного, або один до багатьох
Виміри	Може мати більш ніж одну таблицю вимірів для кожного виміру	Може мати лише одну таблицю вимірів для кожного виміру
Нормалізація	Таблиці вимірів нормалізуються, а таблиці фактів - ні	Обидві – і таблиця фактів, і вимірів – не є нормалізованими
Обсяг таблиці вимірів	Використовується для кубів з великими таблицями вимірів	Використовується для кубів, де обсяг таблиць вимірів є меншим

5. Створення кубів для аналізу даних про ринок нерухомості

Структура бази даних про об'єкти ринку оренди нерухомості України, створена за схемою зірки, представлена на рисунку 1.

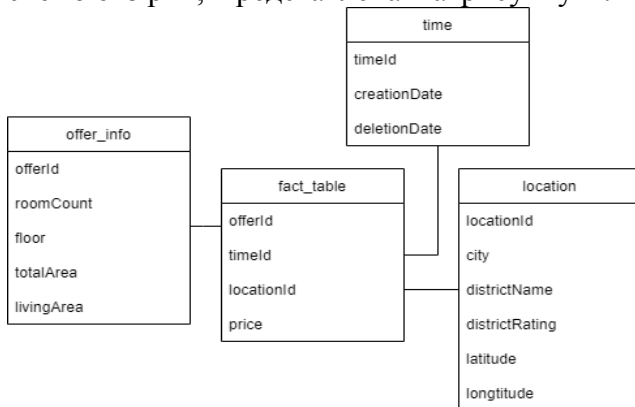


Рис. 1. Структурна схема даних

На її основі побудуємо куби з бажаними вимірами та значеннями для отримання аналітичних даних про ринок нерухомості. Для отримання інформації про зміну середньої ціни об'єкту нерухомості за певний період часу, створено куб, вимірами якого є дата створення оголошення та кількість кімнат, а мірою – середня ціна об'єктів, та

середня ціна за метр квадратний, вимір часу. Аналітичну цінність має спостереження кількості замовлень в залежності від сезону, тому є доцільним використати куб, де вимірами є дата та кількість кімнат, а мірою – число створених заявок за цими параметрами. Для користувачів, що виконують аналіз даних з метою знаходження оптимальної ціни оренди для об'єкту, корисним є здобуття інформації про час перебування об'єкту на дошці оголошень від моменту створення оголошення до його видалення, що, зазвичай, свідчить про факт успішного проведення операції. Тому можемо представити куб з вимірами по кількості кімнат та даті, а мірами буде термін перебування оголошення на дошці та ціна об'єкту. Для проведення аналізу відмінності ціни об'єкту в залежності від його географічного положення, можемо додати куб, де вимірами є місце знаходження, яке описується містом, у якому розташований об'єкт нерухомості, а також районом, кількість кімнат, що є одним з основних формуючих факторів ціни, а мірою задамо ціну об'єкту.

Висновки

Виконано огляд OLAP-технології з метою застосування її для ефективного аналізу значної кількості історичних даних. Описані основні особливості зберігання даних у вигляді OLAP кубів та операції, що можуть виконуватись над кубами. Запропоновано рекомендації щодо створення ефективного та продуктивного кубу. Представлено структуру кубу для отримання аналітичних даних про ринок оренди нерухомості.

Список літератури

1. Pollack, Edward. (2021). OLAP vs. OLTP - How Do They Differ.
2. Queiroz-Sousa, Paulo & Salgado, Ana Carolina. (2020). A Review on OLAP Technologies Applied to Information Networks. ACM Transactions on Knowledge Discovery from Data.
3. Wang, Quankun & You, Jinguo & Zou, Benyuan & Chen, Yu & Huang, Xingrui & Jia, Lianyin. (2021). Reduced Quotient Cube: Maximize Query Answering Capacity in OLAP.
4. CHEN, Zhuo. (2019). An Approach for Developing Platform of OLAP. DEStech Transactions on Computer Science and Engineering.
5. Hassan, Ali & Darmon, Patrice. (2019). Data Reduction in Multifunction OLAP.
6. Marcel, Patrick. (2017). OLAP Personalization and Recommendation. 10.1007/978-1-4899-7993-3_3191-3.