

## ВИКОРИСТАННЯ СУБД VERTICA ДЛЯ ОБРОБКИ ВЕЛИКИХ ДАНИХ

Коливанова Г.В., Яковлева О.В.

*Харківський національний університет радіоелектроніки,  
м. Харків*

В даній роботі проаналізовано питання великих даних, а також інструменти, які існують для зберігання таких даних, наприклад HP Vertica.

Інтерес до технологій Big Data постійно зростає. Популярність цієї технології обумовлена прагненням бізнесу краще розуміти своїх клієнтів та потребою у пошуку нових каналів росту за рахунок відстеження ринкових закономірностей.

Великі дані (Big Data) - це сукупність структурованих і неструктурованих, постійно зростаючих даних, а також методів, інструментів і методів їх обробки в мережі з урахуванням різноманіття форматів і джерел даних. Метою обробки є отримання зрозумілих людині результатів, які узагальнюють поведінку досліджуваних об'єктів і дозволяють її прогнозувати.

На даний момент існує декілька спеціальних платформ і СУБД для зберігання і обробки великих даних: Hitachi Content Platform, Hitachi Network Attached Storage, EMC Isilon, IBM Netezza і т.д.

При виборі сервера зберігання даних для реалізації даної роботи було обрано HP Vertica. HP Vertica – це спеціалізована база даних, яка призначена для аналізу великого об'єму інформації в режимі реального часу. Архітектурні особливості HP Vertica дозволяють обробляти і надавати інформацію набагато швидше і ефективніше, ніж традиційні СУБД. Технології, що дозволяють досягти таких результатів: технологія колоночного зберігання, високоефективне стиснення, масово-паралельна архітектура (MPP), вбудована підтримка високої доступності (HA), стандартний SQL інтерфейс, автоматизований дизайнер БД, ідеологія «нульового» адміністрування.

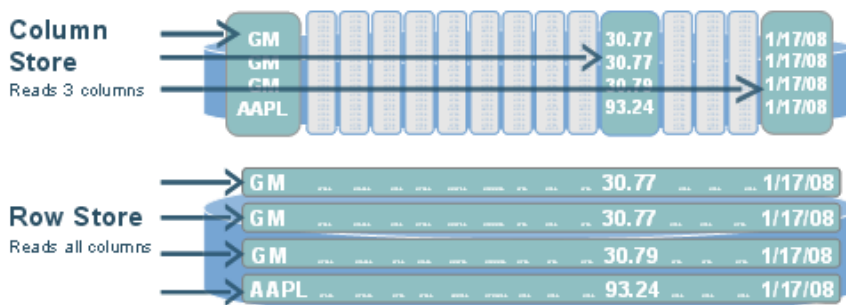


Рис.1 - Зберігання даних в HP Vertica

За допомогою HP Vertica було реалізовано таку задачу - відстеження активності користувачів на сайті в режимі реального часу. Програмно було реалізовано завантаження такої інформації: дані користувача, що ввійшов на сайт, пункти меню, які він обирає, розділи та сторінки, які відкривають для перегляду, а також час, який кожен користувач перебуває на сайті. Ця інформація потрібна для аналітики та покращення сайту в майбутньому.

В питанні великих даних є безліч різних задач, але варто пам'ятати, що для кожного класу задач потрібна своя система, яка допоможе обробляти дані. СУБД є Vertica призначена для вирішення аналітичних задач в реальному часі.