

Технічні науки

УДК 004.658

Демченко Юлія Сергіївна

студентка магістратури

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Демченко Юлия Сергеевна

студентка магистратуры

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Demchenko Yuliia

Master's degree Student of the

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

Якимчук Вікторія Сергіївна

кандидат технічних наук, доцент, доцент кафедри БМК

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Якимчук Виктория Сергеевна

кандидат технических наук, доцент, доцент кафедры БМК

Национальный технический университет Украины

«Киевский политехнический институт имени Игоря Сикорского»

Yakymchuk Viktoriia

Candidate of Technical Sciences, Associate Professor,

Assistant Professor of the BMC's Academic Department

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

**АНАЛІЗ ЗАСОБІВ ДЛЯ РОБОТИ З ДАНИМИ ТА ЇХ
СИНХРОНІЗАЦІЯ
АНАЛИЗ СРЕДСТВ ДЛЯ РАБОТЫ С ДАННЫМИ И ИХ
СИНХРОНИЗАЦИЯ
ANALYSIS OF TOOLS FOR WORKING WITH DATA ITS
SYNCHRONIZATION**

Анотація. Існує безліч середовищ для роботи з базами даних. Що ж, коли прийде час з ними працювати, тоді постане питання вибору. У даній статті було виділено ряд систем для керування базами даних, деякі з них відомі, деякі мені відомі, але в кожній з них можна знайти те, що потрібно саме для обраного проекту.

Ключові слова: база даних, запити, керування, зберігання, дані.

Аннотация. Существует множество сред для работы с базами данных. Что же, когда придет время с ними работать, тогда встанет вопрос выбора. В данной статье был выделен ряд систем для управления базами данных, некоторые из них известны, некоторые менее известные, но в каждой из них можно найти то, что нужно именно для выбранного проекта.

Ключевые слова: база данных, запросы, управление, сохранение, данные.

Summary. There are many environments for working with databases. Well, when the time comes to work with them, then will arise the question of choice. This article highlighted a number of systems for managing databases, some of them are known, some are less well-known, but in each of them you can find what you need for the chosen project.

Key words: database, requests, management, storages, data.

Реляційні бази даних. MySQL є найбільш поширеною та найпопулярнішою системою серед користувачів баз даних та має дочірні системи, такі як MariaDB або Percona. Вказані дочірні системи створені для роботи з простими проектами, які наділені основними перевагами MySQL. База даних MySQL сумісна, практично, з усіма операційними системами, а саме Windows, Linux, Unix, Apple, FreeBSD тощо. Вона зберігає дані в таблицях, що групуються у базу даних, використовує структурну мову запитів (SQL) для доступу до даних та таких команд, як SELECT, UPDATE, INSERT та DELETE для управління ними. База даних MySQL підтримує різні двигуни зберігання даних, такі як InnoDB (за замовчуванням), Federated, MyISAM, Memory, CSV, Archive, Blackhole і Merge. Механізм зберігання даних InnoDB, в основному відрізняється, від інших систем MySQL, наявністю механізму транзакцій та внутрішніх ключів [1]. Формат даних InnoDB забезпечує надійне зберігання даних за рахунок транзакційності та блокування даних на рівні рядку, в чому дана система є набагато кращою, ніж колишній двигун зберігання MyISAM. На сьогоднішній день, MySQL, яка була побудована MySQL AB належить компанії Oracle [2].

PostgreSQL або Postgres також є чудовим вибором для баз даних. MySQL є більш популярною, але головним чином, з історичних причин Postgres – гарний вибір для того, щоб почати вивчати тонкощі мови програмування SQL. Прикладом цього являється те, що вона дозволяє писати користувацькі функції на різних мовах, крім SQL. Postgres також «зробив» кілька кроків на територію NoSQL (не реляційні бази даних). З Hstore (один з доступних модулів PostgreSQL) можна зберігати пари ключ-значення, не покидаючи Postgres, а починаючи з версії 9.4 можна використовувати та працювати з файлами типів JSON та JSONB, що безперечно кидає виклик багатьом систем для роботи з базами даних [3].

TokuDB – це механізм зберігання даних для MySQL, MariaDB або Percona. Він працює як заміна InnoDB, та відповідає тим же критеріям, що і його попередник. TokuDB – це інструмент, для гарного стискання даних. Зазвичай, це економить робочий простір програми в 3 рази, у порівнянні з InnoDB. Однак продуктивність роботи є незмінною, коли набір даних розміщується у оперативній пам'яті. Проте, коли дані стають занадто великими, TokuDB все ще може працювати у прийнятному режимі, у той час, коли продуктивність іншого механізму різко зменшується. Принциповою різницею між TokuDB та InnoDB є те, що перший використовує структуру даних фрактального дерева, тоді як другий – структуру даних B-дерево. Простіше кажучи, якщо дані великі та необхідно використовувати MySQL – необхідно розглянути роботу з вказаним механізмом, що покращить продуктивність [4].

Бази даних NoSQL. MongoDB – це документально-орієнтована база даних, яка не дуже популярна у використанні. Однак ця база даних розпочала поширення завдяки власній швидкості та гнучкості. Для досягнення такої високої швидкості, однією з речей, яку зробили розробники MongoDB, було видалення усіх безпечних записів за замовчуванням. Однією з причин, початку роботи з MongoDB, є те, що його гнучка схема дозволяє вносити зміни «на ходу». Впровадження версії 3.0 значно покращило продуктивність, а також наблизило відповідність баз даних до ACID з новим механізмом зберігання WiredTiger (нова система зберігання даних).

Cassandra – одна з баз даних, спеціально розроблена для обробки великих об'ємів даних. Дослідники з Університету Торонто виявили, що Cassandra була переможцем у досягненні максимальної продуктивності серед всіх експериментів [5]. Вона являє собою гібрид зберігання ключів та табличної бази даних. Таким чином, тут все ще є рядки та стовпці, але

неможливо виконувати об’єднання або підзапити, якщо виконання не проходить за допомогою Hadoop.

ArangoDB – це база даних багатьох режимів, вона організовує дані у вигляді документів, графіків або пар ключ-значення. Завдяки ArangoDB, в ситуації якщо є застосунок, в якому потрібні два способи зберігання даних, не потрібно використовувати кілька баз даних, можна використовувати одну. Для отримання доступу до даних можна використати AQL (мова запитів, яка схожа з SQL). ArangoDB також надає спосіб доступу до себе безпосередньо через REST API. Якщо ви порівняєте ArangoDB з MongoDB, то перша буде більш економічно використовувати пам’ять.

Графічні бази даних. Neo4j – це база даних, спеціально призначена для зберігання всіх своїх даних по вітках і вузлах. Тобто, вона зберігає дані у вигляді графіків. Зберігання даних у такий спосіб дозволяє вирішити деякі задачі набагато швидше. Для даних, які описуються у графіках, можна зробити запит на них в 1000 разів швидше, ніж в реляційній базі даних.

OrientDB – це ще одна графічна база даних, яка дуже схожа до Neo4j. Одне з ключових відмінностей полягає в тому, що OrientDB завжди являється вільна у доступі. Зверніть увагу, що якщо ви хочете корпоративну версію Neo4j, то вона має вільну ліцензію завдяки APGL. Інша ключова відмінність полягає у тому, що OrientDB як і раніше підтримує мову SQL, де у Neo4j є власна мова запитів.

Neo4j є повнофункціональної графічною базою даних, існують й інші більш прості, такі як FlockDB. Ця база даних не намагається вирішити всі задачі з графами. Замість цього вона фокусується на обробці високої пропускної здатності з низькою затримкою. На відміну від Neo4j, вона не виконує глибокого обходу графу.

Висновок. Звичайно, яку базу даних вибрати, повністю залежить від того, яка ціль її використання поставлена перед користувачем. Для проектів в яких важливо використовувати цілісність даних використовуйте

PostgreSQL. Або якщо ж це невеликий додаток - MySQL був би хорошим вибором. У випадку коли пріоритетною вимогою є швидкий час доступу - краще використовувати системи з NoSQL. У Server Density опрацьовується понад 30 ТБ даних на місяць (понад 2 млрд. документів у день), і саме тут MongoDB цінується на вагу золота.

Література

1. MySQL: отличия между MyISAM и InnoDB [Електронний ресурс]. – Режим доступу: <https://itif.ru/otlichiya-myisam-innodb/>
2. MySQL | The Most Popular Open-Source Database | Oracle InnoDB [Електронний ресурс]. – Режим доступу: <https://www.oracle.com/MySQL/>
3. XML, HSTORE, JSON, JSONB—OH MY! [Електронний ресурс]. – Режим доступу: <https://speakerdeck.com/theory/xml-hstore-json-jsonb-oh-my>
4. Benchmarking Percona Server TokuDB vs InnoDB [Електронний ресурс]. – Режим доступу: <http://www.percona.com/blog/2013/05/07/benchmarking-percona-server-tokudb-vs-innodb/>
5. Rabi T., Sadoghi M., Jacobsen H.-A., Gomez-Villamor S., Muntés-Mulero V., Mankovskii S.. Solving Big Data Challenges for Enterprise Application Performance Management (2012) – Proceedings of the VLDB Endowment (PVLDB), Vol. 5, No. 12 – стор. 1724-1735.