ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДІВ ТА ЗАСОБІВ ЗБЕРІГАННЯ І РОЗПОДІЛЕННЯ ДАНИХ У NOSQL СХОВИЩАХ

Омельченко В. О.
Науковий керівник – к.т.н., проф. Лєсна Н. С.
Харківський національний університет радіоелектроніки
(61166, Харків, пр. Науки 14, каф. ПІ, тел. (057)-702-10-13)
Email:vladyslav.omelchenko1@nure.ua

This paper investigates the usage of methods for storage and data distribution. The need of big data research is fast growing industry. The amount of data itself that needs to be stored and analyzed is increasing every day, as well as the number of industries that need to store a large amount of information. The work is devoted to the study of the effectiveness of methods and tools for storing and distributing data in NoSQL storages.

Великі дані (Big Data) - сукупність безупинно збільшуючихся обсягів інформації одного контексту, але різних форматів представлення, а також методів і засобів для ефективної і швидкої обробки [1].

Завдяки експоненціальному зростанню можливостей обчислювальної техніки, описаному в законі Мура [2], обсяг даних не може бути точним критерієм того, чи є вони великими. Наприклад, сьогодні великі дані вимірюються в терабайтах, а завтра - в петабайтах. Тому головною характеристикою Big Data є ступінь їх структурованості і варіантів представлення.

Яскрава ілюстрація великих даних - це безперервна потокова інформація з датчиків або пристроїв аудіо та відеореєстрації, потоки повідомлень з соцмереж, метеорологічні дані, координати геолокації абонентів стільникового зв'язку, тощо.

Програмно-апаратні засоби роботи з Big Data передбачають масштабованість, паралельні обчислення і розподіленість, тому що безперервне збільшення обсягу - це одна з головних характеристик великих даних. До основних технологій відносять нереляційні бази даних (NoSQL)

NoSQL[3] - це підхід до проектування баз даних, який може підтримувати широкий спектр моделей даних, включаючи формати "ключ-значення", "документ", "стовпчик" і "графік". NoSQL, що означає «не тільки SQL», є альтернативою традиційним реляційних баз даних, в яких дані містяться в таблиці, а схема даних ретельно розробляється перед побудовою бази даних. Бази даних NoSQL особливо корисні для роботи з великими наборами розподілених даних.

NoSQL допомагає впоратися з наступними вимогами до обсягу, різноманітності і швидкості великих даних:

- обсяг: підтримка властивостей ACID (атомарность, узгодженість, ізоляція, довговічність) дорога і не завжди необхідна. Іноді ми можемо

мати справу з незначними невідповідностями в наших результатах. Таким чином, ми хочемо мати можливість розділити наші дані на кілька нод;

- різноманітність: одна фіксована модель даних ускладнює включення різних даних. Іноді, коли ми витягуємо дані із зовнішніх джерел, ми не знаємо схему. Крім того, зміна схеми в реляційній базі даних дуже дорога;

- швидкість: дані необхідно не тільки швидко отримувати, але й обробляти і використовувати з більшою швидкістю. Багато типів даних мають обмежений термін зберігання, і їх цінність з часом може знизитися, а в деяких випадках дуже швидко.

Єдиного прийнятого визначення NoSQL не існує, але ось його основні характеристики:

- на відміну від реляційної моделі, NoSQL має досить гнучку схему. Різні рядки можуть мати різні атрибути або структуру. Бази даних часто не розуміють схему. Додатки повинні підтримувати узгодженість схеми, включаючи будь-яку денормалізацію;

- також часто краще справляються із завданнями з дійсно великими даними. Це пов'язано з тим, що бази даних NoSQL слідують підходу BASE (узгодженість даних, доступність або стійкість до поділу) замість ACID;

- у NoSQL узгодженість гарантується тільки через деякий проміжок часу, коли запис зупиняється. Це означає, що запити можуть не побачити найсвіжіші дані. Зазвичай це реалізується шляхом зберігання даних, які записуються в памя’ті;

- нарешті, є поняття, відоме як теорема CAP - виберіть 2 з 3 речей: узгодженість, доступність і допуск на розбиття. Бази даних ACID зазвичай є системами CP, а бази даних BASE - зазвичай AP.

Різноманітність наявних методів та засобів зберігання і розподілення даних у NoSQL сховищах обумовлює актуальність проблеми вибору найбільш ефективних із вказаних методів та засобів. Вирішення цієї проблеми здійснюється за результатами експериментальних досліджень, проведених автором на основі розробленої методики оцінювання і визначених критеріїв ефективності.

Список літератури

1. Sam Madden. From Databases to Big Data – URL: https://ieeexplore.ieee.org/abstract/document/6188576/ (Дата звернення 22.02.2021);

2. R.R. Schaller. Moore's law: past, present and future – URL: https://ieeexplore.ieee.org/abstract/document/591665/ (Дата звернення 22.02.2021);

3. Christof Strauch. NoSql Databases – URL: http://csce.uark.edu/~xintaowu/BDAM/nosqldbs.pdf. (Дата звернення 22.02.2021).