

NoSQL. Альтернатива реляційним системам керування базами даних

Д. А. Меньшиков, Л. В. Тимошенко

В роботі розглядаються нереляційні системи керування базами даних (NoSQL). Необхідність відокремлення від традиційних РСКБД викликана їх обмеженнями в структурі даних що зберігаються. Навідміну від РСКБД, NoSQL рішення дозволяють для окремих записів використовувати довільну структуру та довільні поля. Іншим суттєвим фактором є можливість лінійно масштабувати базу даних для довільної кількості вузлів системи, коли масштабування традиційних РСКБД стає неефективним вже при десятку вузлів. NoSQL надзвичайно перспективна для розподілених та cloud словищ даних.

Вступ

З стрімким розвитком інформаційних систем, розширення сервісів для користувачів та збільшенням клієнтської бази виникла проблема зберігання великих обсягів даних. Традиційні реляційні системи керування базами даних (РСКБД) працюють досить повільно з зростанням кількості даних, вони погано працюють з даними що не представляються табличним видом, наприклад, графи. Тому передові ІТ компанії почали розробку алгоритмів та програмних комплексів для вирішення проблем зберігання та швидкого доступу до даних. Результатом стали так звані NoSQL бази даних.

MapReduce

В 2004 році компанія Google поставила за мету розробити потужну швидку базу даних, яку легко можна масштабувати, контролювати її продуктивність. Так виникла СКБД BigTable. В основі BigTable лежить створений самою Google фреймворк MapReduce, що дозволяє проводити паралельні обчислення над петабайтами даних. При надходженні нових даних головний (master server) “готує”дані, розбиває обчислення на підзадачі і роздає ці задачі підлеглим вузлам (worker node). При зчитуванні даних головний сервер отримує результати роботи підлеглих вузлів, “зклеює” результати і віддає результат. Такий підхід використовує й Apache Foundation в своїй СКБД Hadoop.

Об’єктно-орієнтовані бази даних

Іншим напрямком є об’єктно-орієнтовані бази даних (ООБД). Яскравими представниками цього класу є Db4o, Objectivity/DB, ZODB. Враховуючи, що об’єктно-орієнтований підхід в розробці програмного забезпечення майже витіснив процедурний підхід, немає сенсу зберігати дані в одному форматі, потім створювати об’єкти на базі цих даних конвертуючи їх. Набагато ефективніше створити об’єкт що міститиме ці данні і зберігати увесь час його. Саме таким чином працюють ООБД.

Графові бази даних

В багатьох випадках необхідно працювати з даними що пов’язані один з одним, залежать одне від одного, тобто працювати з даними представленими у вигляді графа. Традиційні РСКБД аж ніяк не можливо ефективно застосувати для роботи з великими об’ємами таких даних. Тому ще одна гілка – графові бази даних (ГБД). В ГБД задаються вузли, відношення між вузлами, їх властивості. Найпопулярнішими ГБД є Neo4j, HyperGraphDB, InfoGrid, AllegroGraph – вони мають гарну продуктивність, легко масштабуються. Один сервер може обслуговувати мільярди об’єктів, а при необхідності можна підключити додаткові сервери. AllegroGraph використовується популярним мікроблогінговим сервісом Twitter.

Key-value бази даних

Найбільш поширеними є Key-value бази даних. Популярність вони набули завдяки системи кешування Memcache розробленої спеціально для сервісу LiveJournal. На сьогоднішній день її використовують в Youtube, Wikipedia Foundation, Facebook, Amazon, SourceForge, Twitter. Дані зберігаються в оперативній пам'яті і доступні по певному ключу. Швидкість memcache в сотні разів перевищує швидкість РСКБД. Інші Key-value БД, наприклад, Redis, SimpleDB, Tokio Cabinet, легко масштабуються та дають можливість записувати дані на диск, що актуально при можливому збої в постачанні електроенергії.

Висновки

Описані СКБД набули значну популярність і довели свою потужність, продуктивність та можливість масштабування. В більшості highload проектах без них просто неможливо обійтись, а простота та надійність не створює проблем при міграції з реляційних систем керування базами даних. Роботи по вдосконаленню та розробці нових NoSQL сховищ триває, а багато компаній вже зараз мігрують на NoSQL рішення.

Список літератури

- [1] Chang, Fay, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, and Robert E. Gruber, Bigtable: A Distributed Storage System for Structured Data. — Google Inc., 2006
- [2] Renzo Angles, Claudio Gutierrez, Survey of graph database models. — ACM Computing Surveys, Feb. 2008
- [3] Kim, Won, Introduction to Object-Oriented Databases. — The MIT Press, 1990

Автори

Дмитро Анатолійович Меньшиков — магістр 1-го року навчання, Фізико-технічний інститут, Національний технічний університет України "Київський політехнічний інститут" Київ, Україна; E-mail: d.menshikov@gmail.com

Луїза Владиславівна Тимошенко — магістр 1-го року навчання, Фізико-технічний інститут, Національний технічний університет України "Київський політехнічний інститут" Київ, Україна; E-mail: luizavladislavna@gmail.com