
빅데이터와 NoSQL 활용방안

박승범* · 이상원** · 안현섭*** · 정인환****

*한국정보화진흥원 경영기획부

**원광대학교 정보전자상거래학부(융복합창의연구소)

***브라운슈바이크공과대학 경영정보학과

****한성대학교 컴퓨터공학과

Application Plan of Big Data and NoSQL

Sungbum Park* · Sangwon Lee** · Hyunsup Ahn*** · In-Hwan Jung****

*Department of Management Planning, National Information Society Agency

**D. of Information & Electronic Commerce (I. of Convergence & Creativity), Wonkwang University

***Department of Wirtschaftsinformatik, Technische Universität Braunschweig

****Department of Computer Engineering, Hansung University

E-mail: parksb@nia.or.kr, sangwonlee@wku.ac.kr, hs.ahn@tu-bs.de, ihjung@hansung.ac.kr

요 약

오랜 기간 동안, 관계형 데이터베이스는 많은 기업에서 널리 사용되어왔다. 데이터베이스의 표준 모형으로서, 데이터 저장과 동시성 제어에서의 뛰어난 영향에도 불구하고, 객체와 관계에서의 불일치에 있어서는 단점이 존재한다. 이러한 배경을 극복하기 위해서, 스키마가 없이도 작동하는 빅 데이터를 위한 새로운 해결책으로 NoSQL이 부각되고 있다. 본 논문에서는 관계형 데이터베이스의 장단점뿐만 아니라, 애플리케이션 데이터베이스와 통합 데이터베이스 간의 비교를 연구하려고 한다. 그리고, 빅데이터를 위한 NoSQL을 정의하고 그 특징을 살펴보겠다.

ABSTRACT

For a long time, Relational Databases have been widely used in many enterprises. Even though the Relational Databases have merits such as excellent power of data preservation, concurrency controlling as a standard model in databases, there are a demerit in the inconsistency of objects and relations. Against this backdrop, NoSQL is being magnified as a new solution for Big Data since it can be operated without any schema. In this paper, we research on not only merits and demerits of Relational Databases but also the comparison between application databases and integrated databases. And then, we define NoSQL for Big Data and check its characteristics.

키워드

Big Data, NoSQL, Application Plan, Integrated Database

I. Introduction

Relational Databases (RDB) have been widely used in many enterprises for a long time. The RDB have merits such as excellent power of data preservation, concurrency controlling as a standard model in databases. But, there are demerits in the inconsistency of objects and relations. Against this backdrop, NoSQL

is being magnified as a new solution for Big Data since it can be operated without any schema. In this paper, we compare application databases with integrated databases. And then, we define NoSQL for Big Data and check its characteristics.

II. Limitations of RDB

Most of enterprises take using RDB for granted. But the databases have both many merits and demerits. Let us check out some merits of RDB as follows. (1) The databases provide more outstanding flexibility than file systems when storing a large volume of data. So, we could find out and collect necessary information in many application programs. (2) Enterprise applications could have many users and generate errors since they are operated concurrently and simultaneously. RDB give a great help in handling these problems by use of concurrency controlling functions. The transaction mechanism could suppress the complexity related to concurrency control. (3) Enterprise applications live in the ecological system where many applications mutually collaborate on many applications. RDB use a method with integrated database sharing in order to store many applications in one database. Even though RDB succeed as a best standard method with these merits, there is a big demerit. The demerit is inconsistency between object and relation as the difference between relational model and data structure within memory. In using databases as an integrating point, there is a try to encapsulate databases in applications. The major factor of database changes is a fact that a large volume of data should be supported in clusters.

III. Application Plan of NoSQL

Against this backdrop, NoSQL begin its existence. NoSQL database provides a mechanism in collecting and retrieving a large number of data that employs less constrained consistency models than RDB. This approach has some motivations such as simplicity of design, horizontal scaling, and finer control over availability. NoSQL are often highly optimized key-value stores. In terms of latency and throughput, with the goal being significant performance benefits, the intentions of NoSQL are simple retrieval and appending operations. In Big Data and real-time web applications, NoSQL databases are used in the fields of significant and growing industry. Allowing SQL-like query languages to be used, NoSQL systems are referred to as “Not only SQL” (Figure 1). The common characteristics of NoSQL are as in the following. (1) NoSQL does not use RDB. (2) NoSQL works well in clusters. (3) NoSQL is an open source. (4) NoSQL was developed and established for the Web environment of the 21st century. (5) NoSQL has no schema. The most important result of No SQL appearance is the durability of redundant stores.

IV. Conclusions

Consequently, NoSQL databases could increase productivity by use of simplification of data access. The major reasons in considering NoSQL are two points. One is data access processing in the condition of the volume of data that clusters need and its performance. The other is to improve development productivity via more convenient operating methods of data.

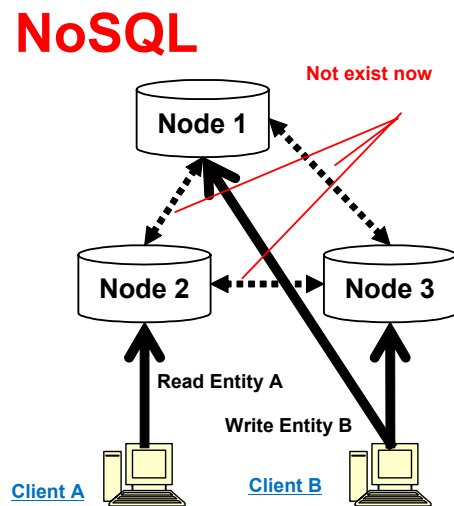
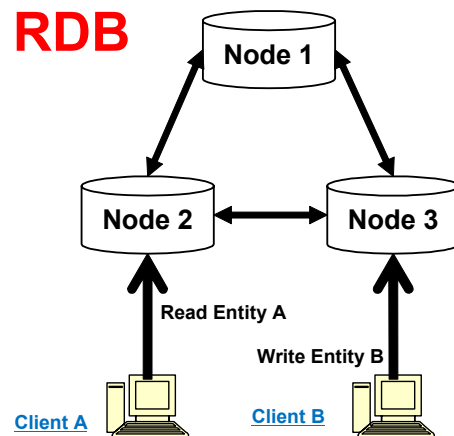


Figure 1. RDB vs. NoSQL

References

- [1] A B M Moniruzzaman and Syed Akhter Hossain, “NoSQL Database: New Era of Databases for Big data Analytics - Classification and Characteristics and Comparison,” International Journal of Database Theory and Application, Vol. 6, No. 4, 2013.
- [2] Burton H. Bloom, “Space/Time Trade-offs in Hash Coding with Allowable Errors,” Communications of the ACM, Vol. 13, No. 1, pp. 422-426, 1970.
- [3] Christof Strauch, NoSQL Databases, Stuttgart Media University Press, 2012.
- [4] Edgar F. Codd, “A Relational Model of Data for Large Shared Data Banks,” Communications of the ACM, Vol. 13, No. 6, pp. 377-387, 1970.
- [5] John Ousterhout, Parag Agrawal, David Erickson, Christos Kozyrakis, Jacob Leverich, David Mazières, Subhasish Mitra, Aravind Narayanan, Guru Parulkar, Mendel Rosenblum, Stephen M. Rumble, Eric Stratmann and Ryan Stutsman, “The Case for RAMClouds:

Scalable High-Performance Storage Entirely in DRAM,” SIGOPS Operating Systems Review, Vol. 43, No. 1, pp. 92-105, 2010.

[6] Joshua Bloch, Effective Java - Programming Language Guide, Addison-Wesley Longman, 2001.

[7] Marcos K. Aguilera, Wojciech Golab and Mehul A. Shah “A Practical Scalable Distributed B-Tree,” Proceedings of the VLDB Endowment, Vol. 2008, No. 1, pp. 598-609, 2008.

[8] Patrick O’ Neil, Edward Cheng, Dieter Gawlick and Elizabeth O’ Neil, “The Logstructured Merge-Tree,” Acta Informatica, Vol. 33, No. 4, pp. 351-385, 1996.

[9] Pramod Sadalage and Martin Fowler, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Addison-Wesley, 2012.

[10] Stavros Harizopoulos, Daniel J. Abadi, Samuel Madden and Michael Stonebraker, “OLTP through the Looking Glass, and What We Found Fhere,” ACM SIGMOD International Conference on Management of Data, Vol. 2008, No.1, pp. 981-992, 2008.