

Chapter 4

An Empirical Study of NoSQL Databases for Big Data

Wen-Chen Hu

University of North Dakota, USA

Hongyu Guo

University of Houston – Victoria, USA

Naima Kaabouch

University of North Dakota, USA

Hung-Jen Yang

National Kaohsiung Normal University, Taiwan

ABSTRACT

Relational databases have dominated the database markets for decades because they perform extremely well for traditional applications like electronic commerce and inventory systems. However, the relational databases do not suit some of the contemporary applications such as big data and cloud computing well because of various reasons like their low scalability and unable to handle a high volume of data. NoSQL (not only SQL) databases are part of the solution for developing those newer applications. The approach they use is different from the one used by relational databases. This chapter discusses NoSQL databases by using an empirical instead of theoretical approach. Other than introducing the types and features of generic NoSQL databases, practical NoSQL database programming and usage are shown by using MongoDB, a NoSQL database. A summary of this research is given at the end of this chapter.

INTRODUCTION

Much of current data processing requires horizontal scaling, faster speed, and processing different kinds of data and relational databases are not able to meet the requirements adequately. A NoSQL database is to meet the needs by providing a simple and efficient mechanism for data storage and retrieval. The approach it uses is different from the one used by a relational database. An RDBMS (relational database management system) is a general-purpose data store, whereas a NoSQL database is normally a key-value store for simple insertion and retrieval operations. This chapter discusses the NoSQL databases by using an empirical approach. A NoSQL database, MongoDB, is used to show how to program and use NoSQL databases.

DOI: 10.4018/978-1-5225-0182-4.ch004

Importance of Big Data

In the information age, data is generated explosively. For example, everyday there are 230 million tweets sent, 294 billion emails delivered, and 100 terabytes of data uploaded to Facebook according to IDC (2012). Important and useful information such as crime prevention and customer buying patterns could be buried inside the huge amount of unstructured data. How to effectively manage and use the big data is a big headache for organizations. Many IT researchers and scholars and organizations start working on this problem. Even the Obama Administration (The White House, 2012) announced a “Big Data Research and Development Initiative.” Several US Federal departments and agencies have pledged more than \$200 million for this big data initiative. The departments and agencies and their big data themes include

- Together National Science Foundation (NSF) and the National Institutes of Health (NIH): *Core Techniques and Technologies for Advancing Big Data Science & Engineering*,
- National Science Foundation (NSF): Including funding the big data solicitations, keeping with its focus on basic research, and implementing a comprehensive, long-term big data strategy,
- Department of Defense (DOD): *Data to Decisions*,
- National Institutes of Health (NIH): *1000 Genomes Project Data Available on Cloud*,
- Department of Energy (DOE): *Scientific Discovery Through Advanced Computing*, and
- US Geological Survey (USGS): *Big Data for Earth System Science*.

Big data is everywhere now and valuable information is hiding in it. In the past, this information was simply ignored and opportunities were missed. Realizing the great importance of big data, organizations scramble to find hidden information buried in big data and try to make the best use of it.

Features of NoSQL Databases

Many of contemporary applications require the features of being generic, distributed, open-source, and scalable. The features are also some of the benefits of NoSQL databases including (Han, E, Le, & Du, 2011):

- NoSQL databases are able to handle structured, semi-structured, and unstructured data, not just semi-structured data for relational databases.
- Compared to relational databases, NoSQL databases are simple and therefore are faster.
- Most NoSQL databases allow object-oriented programming, which is the most popular programming paradigm.
- NoSQL databases include the feature of horizontal scalability, which is the ability to connect multiple NoSQL databases so that they work as a single logical unit.

Relational databases can no longer meet the requirements of some of the mobile-age applications, which desire simple, efficient, and scalable databases. The emergence of NoSQL databases is to meet the needs of newer applications. However, from the history of databases, none of other databases has successfully replaced relational databases, which have the greatest advantages of easy-to-use. NoSQL databases do not intend to replace relational databases. Instead, they are a great complement to relational databases.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/an-empirical-study-of-nosql-databases-for-big-data/157684

Related Content

A Route Recommender System Based on Current and Historical Crowdsourcing

Marlene Goncalves, Patrick Rengifo, Daniela Andreina Rodríguez and Ivette C. Martínez (2017). *Social Media Data Extraction and Content Analysis* (pp. 114-136).

www.irma-international.org/chapter/a-route-recommender-system-based-on-current-and-historical-crowdsourcing/161962

Deep Learning-Based Adaptive Online Intelligent Framework for a Blockchain Application in Risk Control of Asset Securitization

Liuyang Zhao, Yezhou Sha, Kaiwen Zhang and Jiaxin Yang (2023). *International Journal of Data Warehousing and Mining* (pp. 1-21).

www.irma-international.org/article/deep-learning-based-adaptive-online-intelligent-framework-for-a-blockchain-application-in-risk-control-of-asset-securitization/323182

Handling Local Patterns in Collaborative Structuring

Ingo Mierswa, Katharina Morik and Michael Wurst (2008). *Successes and New Directions in Data Mining* (pp. 167-186).

www.irma-international.org/chapter/handling-local-patterns-collaborative-structuring/29959

Using TreeNet to Cross-sell Home Loans to Credit Card Holders

Dan Steinberg, Nicholas C. Cardell, John Ries and Mykhaylo Golovnya (2008). *International Journal of Data Warehousing and Mining* (pp. 32-45).

www.irma-international.org/article/using-treenet-cross-sell-home/1805

Multidimensional Model-Based Decision Rules Mining

Qinrong Feng, Duoqian Miao and Ruizhi Wang (2009). *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction* (pp. 311-334).

www.irma-international.org/chapter/multidimensional-model-based-decision-rules/8449