Int. j. inf. tecnol.
https://doi.org/10.1007/s41870-020-00515-8

ORIGINAL RESEARCH

Schema design advisor model for RDBMS to NoSQL database migration

Basant Namdeo[1] · **Ugrasen Suman**[2]

Received: 11 February 2020 / Accepted: 11 September 2020

Abstract Database schema design has a significant importance in software design. There are lots of tools and methods available for schema design in RDBMS but limited attention is given in NoSQL for schema design as it is emerging in database technology. NoSQL requires a different approach in designing efficient schema, such as in the document database which information should be stored as embedded document, or which information should be stored as referenced document. There are certain thumb rules for schema design in NoSQL databases. In reengineering projects, especially in Old RDBMS to new NoSQL system, developing good and efficient database schema is a very difficult task. In this paper, we have proposed a schema design advisor model which uses the existing software's SQL queries load as an input along with an algorithm for schema design recommendation. Also, we have proposed a cost model for various schemas created by recommendation model. The proposed model is implemented through a prototype for the MongoDB document database in Java. The prototype produces all possible combinations of schemas and calculates cost of each schema. Automated schema design process produces all possible combinations of NoSQL schemas, which is difficult with manual schema design approach.

✉ Basant Namdeo
basant_nd@yahoo.com

Ugrasen Suman
ugrasen123@yahoo.com

[1] International Institute of Professional Studies, Devi Ahilya University, Indore, India

[2] School of Computer Science and IT, Devi Ahilya University, Indore, India

Keywords Database reengineering · Schema design model · NoSQL cost model · RDBMS

1 Introduction

In big data era, NoSQL database system has become the first choice of database designers because of its high availability, performance and scalability. Relational databases follow the ACID properties for accuracy, completeness and data integrity in database. On the other hand, NoSQL databases relax on ACID properties, and instead these follow the BASE (Basically Available, Soft state and Eventually consistence) model. In distributed systems, three properties are commonly desired – namely, consistency, availability, and partition tolerance [1]. Also it's impossible to achieve all three at a time in a system. Relational databases such as MySQL, SQL Server follows Consistency and Availability (CA), and NoSQL database such as MongoDB, Redis etc., follows Consistency and Partition tolerance (CP) [2].

Database schema design plays an important role in database centric application development. RDBMS can only stores data, which are in a fixed format. But in today's era, in which lots of new devices (IoT based) has come that they produce lots of data regularly, and this data generally not follows any fixed format. The social media websites such as Facebook, Twitter, WhatsApp, LinkedIn and Hike messenger, providing the users to express their feelings also by using the different symbols like smiley's, funny faces, and various gif animation files etc.[3]. For storing and manipulating this vast and heterogeneous data, companies require a new type of database technology. Therefore, NoSQL has become an important and useful database in order to solve such problems. NoSQL is a term, which is

Published online: 20 September 2020