

NoSQL Database Design Using UML Conceptual Data Model Based on Peter Chen's Framework

Kwangchul Shin, Chulhyun Hwang, Hoekyung Jung*

*Department of Business IT, Kook Min University, 77 Jeongneung-Ro, Seongbuk-gu, Seoul, South Korea,
Department of Computer Engineering, Pai Chai University, Doma2-Dong,
Seo-gu, Daejeon, South Korea.
Correspondent Author: Hoekyung Jung*

Abstract

In the Big Data era, relational databases and NoSQL databases coexist in Polyglot Persistence environment. Although data management is more essential in an environment where a variety of databases are, NoSQL databases only concentrate on solving non-functional requirements to run well on large clusters. This situation makes consistent data management standards difficult. To solve this problem, this study proposes NoSQL database design method using conceptual data model based on Peter Chen's framework. The proposed design method is applied to the e-commerce business area in order to examine the applicability of it.

Keywords: Data Model, Conceptual Data Model, Big Data, NoSQL, Database design

INTRODUCTION

In the Big Data era, data architecture, which had been mainly made of relational databases for decades, has currently both relational databases and NoSQL databases in Polyglot Persistence environment. Being studied for a long time, relational database has design methods to implement database from data requirements but NoSQL database design lacks researches on the design methods. This is because NoSQL databases only concentrate on solving specific application problems of non-functional requirements like performance, availability, scalability and so on. When two types of databases which differ in design approaches coexist, the difference between them becomes high risks in data management aspect. Therefore, data perspective should be considered in NoSQL database design [14].

This work raises the problem of no conceptual design phase in NoSQL database design and proposes applying conceptual data modeling, which, is mainly used at relational database design, to NoSQL database design based on Peter Chen's suggestion to solve the problem. Data management in a consistent method is required in various types of databases after the advent of NoSQL. Accordingly, conceptual data model gets more important, which can be expressed as a consistent information schema without dependence of specific databases.

The present paper is composed of five chapters: chapter 2 addresses the necessity of conceptual data model applied to NoSQL database design through related researches; chapter 3 explains the method to design NoSQL database using conceptual data model; chapter 4 illustrates the real examples applied, to find out the possibility of using the proposed data modelling method in Document Database; and chapter 5 concludes the paper with the proposal on the future work plan.

RELATED WORKS

Conceptual data model which can express only data in itself independent of particular technologies has been variously utilized for data sharing, reference data model and so on [10,11]. In database design, conceptual data model does not depend on a specific data model like relational, network and hierarchy [2]. Therefore, it can also apply NoSQL database design. Even though there is a study on how to transform conceptual data model into logical data models which are relational model, network model and hierarchy model [8], a study on methods to change conceptual data model into NoSQL database lacks.

Except for Graph, NoSQL data model like Key Value, Document and Column Family represents aggregate data model [15]. Aggregate data model, the target for NoSQL database design, is a chunk of related data, a unit of data access, manipulation and distribution (i.e., sharing) [7]. Studies on NoSQL database design are conducted in two directions: one is the design method that specialize in specific NoSQL data model by approaches to each NoSQL data model [5, 13], and the other is the design method independent of specific NoSQL data model by defining logical aggregate data model [4, 12]. But studies on the design method using conceptual data model lacks.

Existing studies don't have had not only NoSQL database design using conceptual data model but also transformation from conceptual data model to NoSQL data model. In the next chapter, the proposal is foundation and the method of applying conceptual data model to NoSQL database design.

NoSQL DATABASE DESIGN

NoSQL Data Model :

The proposal is how to use conceptual data model in NoSQL database design based on Peter Chen's framework. Data model used in NoSQL database as aggregate data model represents data set in which specific applications are interested and means Access Path dependent data structure. In his framework, logical views of data are identified as four levels like below [6]. NoSQL data model is mainly concerned with levels 3 and 4.

1. Information concerning entities and relationships which exist in our minds.
2. Information structure-organization of information in which entities and relationships are represented by data.

3. Access-path-independent data structure-the data structures which are not involved with search schemes, indexing schemes, etc.
4. Access-path-dependent data structure.

Figure 1 below which is his changed framework represents levels of views of data for entity relationship model, relational model and NoSQL model. In levels 2, entity relationship diagram is similar to 3NF relation and the result of entity relationship modeling can be easily transformed into relation model. The result set written by SQL on relational model is similar to NoSQL data model which is aggregate data model. Therefore, entity relationship diagram and relational model can be used in NoSQL database design.

Levels of Logical Views		Data Models		
		Entity Relationship	Relational	NoSQL
Level 1	Information Concerning Entities and Relationships	Entities, Relationships, Attributes		
Level 2	Information Structure	Entity-Relationship Diagram	3NF Relations <i>Decomposition Approach</i>	
Level 3	Access Path Independent Data Structure	Table	Relations(Tables)	
Level 4	Access Path Dependent Data Structure			Key-Value , Document , Colum Family , Graph

Figure 1. Logical view of NoSQL Data Model

Phases of NoSQL Database Design :

Phases of proposed NoSQL database design consists of conceptual, logical and physical design in 3steps like phases of traditional relational database. When designing relational database, conceptual data model is represented as relational data model, a kind of logical data model. A specific DBMS like Oracle, SQL server and MySQL is implemented from the relational data model. NoSQL database design also can have the same 3 phases like relational database that is drawing conceptual data model, transforming into logical data models (Key-Value, Document and Column Family) and building up specific NoSQL databases. The detailed description for each phases is followings:

Conceptual Design: Conceptual data modeling is the process developing conceptual schema of database from user's requirements. Conceptual schema which is independent of a NoSQL database to implement describes high level of database structure. The purpose of conceptual data modeling is for the content included in information rather than for data storage

structure. Conceptual data modeling can be applied not only to NoSQL database but also to relational database.

Logical Design: Logical data modeling is the process developing logical schema of NoSQL database from conceptual schema. Logical schema describes data structures managed in NoSQL database. It is dependent on data models used in NoSQL database but independent from NoSQL database software. In other words, all document databases like Mongo DB, Couch DB and so on in logical design phase are represented as the same one document data model and so we can choose document database software lately after the document data modeling.

Physical Design: Physical data modeling is the process developing physical schema of NoSQL database. Physical schema describes data structure implemented in a certain NoSQL database software; it is a description of storage structures and effective methods for data access. Therefore, physical design has an application of features of a specific NoSQL database software to physical schema in order to improve performance and availability.

Table 1 shows the dependence of conceptual, logical, and physical design on the NoSQL data model and specific NoSQL database.

Table 1. Dependence of conceptual, logical, and physical design on the NoSQL data model and specific NoSQL database

Dependence of on:	NoSQL Data Model	Specific NoSQL Database
Conceptual Design	No	No
Logical Design	Yes	No
Physical Design	Yes	Yes

How to use conceptual data model for NoSQL database design :

Conceptual data model is conceptual schema which represents real world by data and high level data model independent from specific logical data models and databases. ER, UML, ORM and FCO-IM can be used for it [1, 2, 6, 9], which is unlimited for application of NoSQL data model and NoSQL database. Table 2 shows applicable targets of conceptual data model, NoSQL data model and NoSQL database in NoSQL database phases.

Table 2. Targets in phases of NoSQL database design

Phases	Targets		
Conceptual Design	Conceptual Data Model	ERD(Entity Relationship Diagram) UML(Unified Modeling Language) ORM(Object Role Modeling) FCO-IM(Fully Communication Oriented Information Modeling)	
Logical Design	NoSQL Data Model	Key-Value , Document, Column Family , Graph	
Physical Design	NoSQL Database	Key-Value	Riak, Redis, Memcached ,Berkeley DB ,Hamster DB, Amazon Dynamo DB ,Project Voldemort
		Document	MongoDB, Couch DB, Terrastore, Orient DB, Raven DB
		Column Family	Cassandra, HBase, Hypertable, Amazon Simple DB
		Graph	Neo4J, Infinite Graph, Orient DB, Flock DB

For example, when designing Mongo DB, which is leading document database, conceptual data model independent from specific NoSQL data model can be made using ER, UML, ORM and FCO-IM. After the generated conceptual data model transforming into logical document data model, Mongo DB which is the target database is implemented from the document data model.

DESIGNING DOCUMENT DATABASE USING CONCEPTUAL DATA MODEL

Description of the conceptual data model :

The conceptual data model used in document database design is Figure 2 below, which is a UML data model used as an example in the [3]. In the [3], the UML data model is mentioned as logical data model but it is utilized as conceptual data model in this article because it is not dependent on specific data models and databases. The conceptual data model used for example is about e-commerce business in which a customer orders products based on internet or mobile.

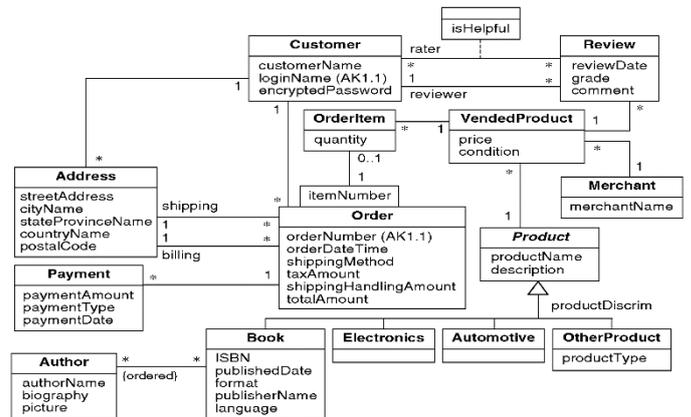


Figure 2. UML conceptual data model

A Customer can place many Orders and an Order belongs to a Customer. An Order may have a shipping address and a billing address differently. An Order can be done by multiple Payments which can be by a credit card, a check and money

An Order may consist of many OrderItems, each of which is identified by an itemNumber within an Order. Each OrderItem has more than one quantity. An OrderItem pertains to a VendedProduct which is a Product sold by a particular Merchant. The price of a Product can vary by Merchant and condition such as new, good condition, and worn.

Products for sales is three kinds: Book, Electronics, and Automotive. Many kinds of Product can be added for sales. Customers can not only place Orders, but also submit Reviews for VendedProducts.

How to transform conceptual data model into document data model :

The components of UML conceptual data model can be switched into the components of document data model like Table 3 below. Class can be transformed collection, set of

documents. Attribute can be mapped to column. Association can be mapped to reference or embedded by application's query patterns.

Table 3. Mapping between components of UML Class Diagram and Document Data Model

UML Class Diagram	Document Data Model
Class	Collection
Attribute	Column in Document
Association	References of Embedded

NoSQL databases supporting document data model can add and alter columns by application's requirements without predefined columns. But if there are data items which already are known, the promise is for defining columns first to implement target database.

Because UML class diagram does not support primary key and foreign key important to implement database, the advantage of

it in conceptual data modeling is being focused on business requirements and scope on the other hand, the disadvantage of it in logical data modeling is not being focused on data structure of document database by itself. To solve this disadvantage by using UML's stereotype like followings: An identifier, which consists of more than one attribute of a collection, is marked to “<<PK>>” in front of the attributes as primary key. An association, which consists of the same attributes between different collections, is marked to “<<FK>>” in front of the attributes as foreign key and to “<<reference>>” or “<<embedded>>” on the line.

Transformed document data model from conceptual data model :

Figure 3 below is the result of transforming UML conceptual data model into document data model. UML class diagram including stereotype is used for representation of document data model.

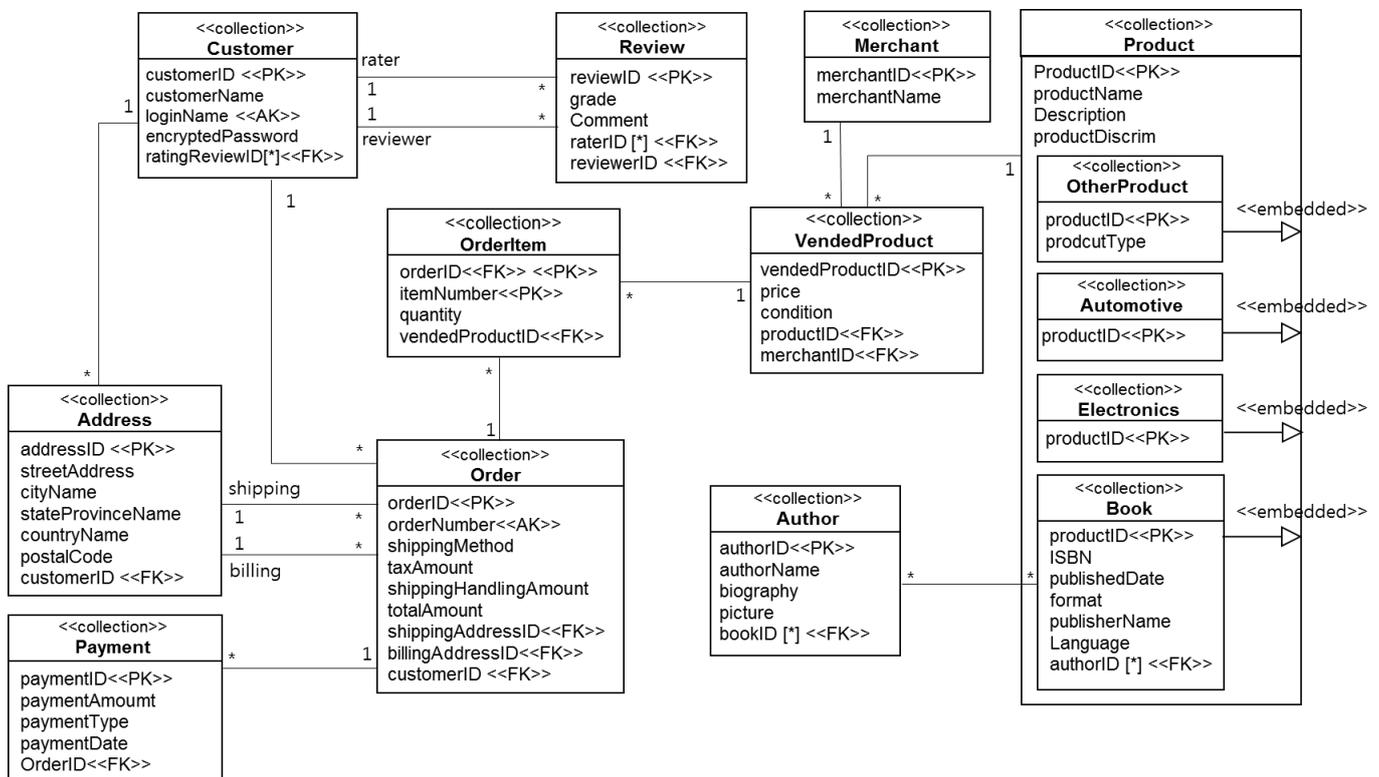


Figure 3. Document Data Model using UML Class Diagram

This article only covers from conceptual phase to logical phase. Inputs for physical design, query patterns which are determined by application characteristics are omitted. So, all associations, which are “<<reference>>” or “<<embedded>>” that depends on query patterns, are represented into “<<reference>>” excluding Product's inheritance relationship marked into “<<embedded>>” in figure 3 above.

CONCLUSION AND FUTURE WORK AREAS

The proposed NoSQL database design method is useful for polyglot persistence with various databases in the big data era and has the following advantages: First, data quality and standard in an enterprise perspective are ensured from conceptual design phase through the conceptual data model. Second, even if implemented in various databases including

NoSQL databases, data management focusing on the conceptual data model can be consistent and integrated. The future studies will be how to transform conceptual data model into diverse NoSQL data model and data management method for NoSQL database.

This work was supported by the research grant of Pai Chai University in 2016.

REFERENCES

- [1] F. N. Azizah, G. P. Bakema, B. Sitohang, and O. S. Santoso, 2009, "Generic Data Model Patterns using Fully Communication Oriented Information Modeling (FCO-IM)." *Electrical Engineering and Informatics*, 2009. ICEEI'09. International Conference on. Vol. 2. IEEE.
- [2] Batini, Carlo, Stefano Ceri, and Shamkant B. Navathe, 1991, "Conceptual database design: an Entity-relationship approach."
- [3] Blaha, Michael, 2013, "UML Database Modeling Workbook." Technics Publications.
- [4] F. Bugiotti, L. Cabibbo, P. Atzeni, and R. Torlone, 2013, "A Logical Approach to NoSQL Databases."
- [5] Chebotko, Artem, Andrey Kashlev, and Shiyong Lu, 2015, "A Big Data Modeling Methodology for Apache Cassandra." *Big Data (BigData Congress)*, 2015 IEEE International Congress on. IEEE.
- [6] Chen, Peter Pin-Shan, 1976, "The entity-relationship model—toward a unified view of data." *ACM Transactions on Database Systems (TODS)* 1.1, pp 9-36.
- [7] E. Evans, 2003, "Domain-Driven Design: Tackling Complexity in the Heart of Software." Addison-Wesley.
- [8] Fahrner, Christian, and Gottfried Vossen, 1995, "A survey of database design transformations based on the Entity-Relationship model." *Data & Knowledge Engineering* 15.3, pp 213-250.
- [9] Halpin, Terry, and Tony Morgan, 2010, "Information modeling and relational databases." Morgan Kaufmann.
- [10] C. H. Hwang, H. Han, K. C. Shin, and H. K. Jung, 2014, "Study on Developing Reference Data Model to Share and Open Data in Public Domain." *International Journal of Software Engineering and Its Applications* 8.12, pp 117-124.
- [11] C. H. Hwang, I. J. Son, and H. K. Jung, 2014, "A Study on Data Modeling Approach Utilizing 3-tier Architecture of 'BEE (Being-Experience-Event)'." *International Journal of Control and Automation* 7.5, pp 35-40.
- [12] Jovanovic, Vladan, and Steven Benson, 2013, "AGGREGATE DATA MODELING STYLE." *SAIS 2013 Proceedings*, pp 70-75.
- [13] Kaur, Kanwalpreet, and Rinkle Rani, 2013, "Modeling and querying data in NoSQL databases." *Big Data*, 2013 IEEE International Conference on. IEEE.
- [14] M. Naab, S. Braun, T. Lenhart, S. Hess, A. Eitel, D. Magin, and F. Kiefer, 2015, "Why Data needs more Attention in Architecture Design-Experiences from prototyping a large-scale mobile app ecosystem." *Software Architecture (WICSA)*, 2015 12th Working IEEE/IFIP Conference on. IEEE, pp. 75~84.
- [15] Sadalage, J. Pramod, and Martin Fowler, 2012, "NoSQL distilled: a brief guide to the emerging world of polyglot persistence." Pearson Education.