

Modelagem Lógica para Bancos de Dados NoSQL: uma revisão sistemática

Victor Martins de Sousa¹, Luis Mariano del Val Cura²

¹Instituto Federal Catarinense – Campus Avançado Sombrio - (IFC - Sombrio)
Sombrio – SC – Brasil

²Faculdade Campo Limpo Paulista (FACCAMP) – Campo Limpo Paulista – SC - Brasil

`victor.sousa@sombrio.ifc.edu.br, delval@faccamp.br`

Abstract. *Cloud Computing is very used nowadays for the storage of high volumes of data in several separate servers. NoSQL databases attends this need, supporting the horizontal scalability and manage high volumes of data. NoSQL databases are divided in four types of data: Key-Value, Oriented-columns, Oriented-Documents and Oriented-Graphs. The project of a BD, influences directly their performance and consistency during use. The logic modeling, one step of the project is reviewed systematically. It was concludes an absence of a motion for modeling logic that can be used in any of the four types of BDs NoSQL.*

Resumo. *Cloud Computing é muito utilizada atualmente, pelo armazenamento de grande volume de dados em diversos servidores separados. Essas necessidades são supridas pelos Bancos de Dados NoSQL, que suportam a escalabilidade horizontal e gerenciam grandes volumes de dados. BDs NoSQL são divididos em quatro modelos de dados: Chave-Valor, Orientado a Colunas, Orientado a Documentos e Orientado a Grafos. O projeto de um BD influencia diretamente no seu desempenho e na sua consistência durante a utilização. A modelagem lógica, uma das etapas do projeto, é revisada de maneira sistemática. Conclui-se a ausência de uma proposta de modelagem lógica que possa ser utilizada em qualquer um dos quatro tipos de BDs NoSQL.*

1. Introdução

A quantidade de dados armazenados na web vem aumentando cada vez mais, fatores como o *Big Data* e as redes sociais têm relação direta com esse aumento. Esses fatores normalmente utilizam o conceito de *Cloud Computing*, que consiste no armazenamento dos dados em diversos servidores separados. Esse tipo de armazenamento não oferece grande desempenho, quando é utilizado nos tradicionais Bancos de Dados (BD) Relacionais, que armazenam os dados estruturados de maneira rígida em tabelas. Com essa nova necessidade, surgiu o conceito NoSQL (*Not only SQL*) com uma estrutura de simples gerência para grandes volumes de dados e principalmente, apoiando escalabilidade, que é uma das principais necessidades das aplicações atuais. ([Bugiotti 2014],[Lima 2015])

Os BDs NoSQL são classificados em quatro modelos: (i) Chave-Valor: possui o conceito de ser uma grande tabela *hash*, onde sua estrutura é baseada em um conjunto de chaves, e cada chave referencia um valor; (ii) Orientado a Colunas: nesse modelo os dados são identificados por colunas e também, podem ser agrupados em dados do mesmo tipo, formando famílias de colunas; (iii) Orientado a Documentos: consiste em um modelo contendo uma coleção de documentos, onde cada documento armazena campos e valores; (iv) Orientado a Grafos: os dados são armazenados em nós de um grafo e possui relacionamentos entre os nós quando existe um tipo de associação [Souza 2014].

O projeto de um BD é necessário para alcançar eficiência, escalabilidade e outras características para obter o melhor desempenho na sua utilização. Esse projeto é separado em três etapas; modelagem conceitual, onde os dados são relacionados de uma forma abstrata; modelagem lógica, onde os dados são modelados para algo mais próximo da utilização; e por fim, a modelagem física, que é o BD implementado [Esmalri, 2011].

Segundo Bugiotti (2014), não existem muitos trabalhos que abordem a conversão do modelo conceitual para o modelo lógico. E também, não existem ferramentas que definam um projeto lógico para os principais modelos de BDs NoSQL.

Este artigo tem por objetivo realizar uma revisão sistemática das ferramentas, ou metodologias, para modelagem lógica de BDs NoSQL. A seção 2 apresenta a metodologia utilizada para a revisão sistemática, a seção 3 apresenta os trabalhos encontrados sobre o tema, a seção 4 apresenta uma breve comparação entre os trabalhos encontrados e a seção 5 apresenta as considerações finais.

2. Materiais e métodos

Este artigo é uma revisão sistemática baseada na proposta de Kitchenham et. al. (2009). O objetivo desta revisão é encontrar contribuições para modelagem lógica, ou projeto, de BDs NoSQL. Alguns trabalhos que contribuem de forma indireta também foram considerados, que não são propostas para modelagem lógica NoSQL, mas abordam conceitos relevantes a essa pesquisa.

As buscas foram realizadas nos engenhos de busca *IEEE Xplorer*, *ACM Digital Library* e *CiteSeerX*, no mês de Junho de 2016. A expressão utilizada na busca foi: *("NoSQL") AND ("column-oriented" OR "key-value" OR "document-oriented" OR "graph-oriented") AND ("conceptual modeling" OR "guidelines" OR "best practices" OR "database design" OR "data modeling" OR "logic modeling")*.

Nas pesquisas foram encontrados 244 artigos. Após análise do resumo e conclusão de cada trabalho, foram selecionados 12 que contém alguma contribuição para a modelagem lógica de BDs NoSQL. Os artigos que serão apresentados de maneira sucinta na seção seguinte, e também apresentados na tabela 1 da seção 4.

3. Trabalhos encontrados

O artigo [Banerjee 2015] apresenta um projeto de BD NoSQL para utilização em Big Data, o projeto consiste em utilizar modelo conceitual que é proposto no artigo,

chamado GOOSSDM (*Graph Object Oriented Semi-Structured Data Model*). O modelo lógico é representado através da utilização do JSON (*Java Script Object Notation*) e a implementação é feita em um BD NoSQL Orientado a Documento.

O trabalho [Bansel 2016] propõe a migração entre modelos de BD NoSQL. As migrações são: Orientado a Documentos para Orientado a Grafos; Orientado a Coluna para Orientado a Grafos e Orientado a Documento para Orientado a Colunas. Essa última migração é feita utilizando um metamodelo, como um intermediário para eliminar as diferenças entre os modelos.

O artigo [Bermbach 2015] apresenta uma abordagem sistemática de um projeto para BD NoSQL Orientado a Colunas. Utiliza-se a criação de um esquema de forma automatizada que é gerado para aplicação específica, através de uma classificação. Inicialmente são levantadas todas as opções viáveis para responder um conjunto de consultas. Em seguida, é realizada uma otimização, que consiste em remover ou realizar fusão de tabelas excedentes. Posteriormente, realiza-se uma fase de classificação, o esquema é definido através de uma ordem de pontuação.

O trabalho [Bugiotti 2014] apresenta uma proposta para projeto de BDs NoSQL. É apresentado um modelo de dados abstrato, que é chamado de NoAM (*Model Abstract NoSQL*). O NoAM fornece uma representação intermediária para os BDs NoSQL, independente do sistema, através da utilização das semelhanças e abstração das diferenças em cada modelo NoSQL. O artigo baseia-se em algumas atividades: modelagem conceitual, identificar entidades e relacionamentos; projeto agregado, agrupar entidades relacionadas em agregados; particionamento agregado, dividir agregados em elementos menores; projeto do BD NoSQL em alto nível, mapear uma representação intermediária através da utilização do NoAM; e por fim, a implementação, mapeia a modelagem intermediária para a modelagem do sistema específico.

O artigo [Feng 2015] propõe a transformação de diagramas de classes UML em um modelo de dados para BD NoSQL Orientado a Colunas. É utilizado um metamodelo para transformação em UML, que é utilizado para fazer uma simplificação. Um metamodelo também é utilizado para a conversão para BD Orientado a Colunas. O trabalho utiliza um sistema para geração de anotações, que trazem informações extras sobre o esquema gerado.

O trabalho [Fioravanti 2016] apresenta uma comparação de resultados de desempenhos obtidos entre um BD Relacional e alguns BDs NoSQL, que utilizaram uma combinação lógica de domínio baseada em padrões com uma camada de persistência. A camada de persistência baseada em um BD Relacional + Java *Persistence API* (JPA) é convertida em novos modelos: NoSQL Orientado a Documentos, que apresentou melhorias em desempenho, e NoSQL Orientado a Grafos, que apresentou uma conversão e implementação simples.

O trabalho [Karnitis 2015], apresenta uma migração de BD Relacional para um BD NoSQL Orientado a Documentos. O processo de migração utiliza dois níveis lógicos para os dados. O primeiro nível é um modelo abrangente e o segundo nível é mais refinado, pois é definido conforme o modelo de documento necessário.

O artigo [Kaur 2013] apresenta um estudo de caso utilizando a modelagem de dados e a sintaxe de consulta para BD relacionais e BD NoSQL Orientados a Grafos e Orientados a Documentos. A modelagem é feita de maneira única para cada tipo de BD: Modelo E-R para os BDs Relacionais, o diagrama de classe é utilizado para BD Orientados a Documentos e uma ferramenta específica é utilizado para BD Orientado a Grafos.

O trabalho [Lima 2015] propõe um projeto lógico para BDs NoSQL Orientados a Documentos. Na proposta, o modelo conceitual consiste em um esquema de *Extended Entity-Relationship* (EER), ou Entidade-Relacionamento Extendido, que possui adaptações que consideram a carga de trabalho esperada na utilização da aplicação. Também é proposto um modelo lógico de Documento NoSQL, a sua conversão é realizada através de um algoritmo apresentado no trabalho. Um estudo de caso é apresentado avaliando duas situações; um teste utilizando projeto convencional e outro teste, com o projeto otimizado, que considera a carga de trabalho no modelo conceitual. No estudo de caso é realizada a implementação em BD NoSQL de Documento, que apresenta melhor desempenho utilizando o projeto otimizado, avaliando a carga de trabalho no início, quando comparado a um projeto convencional.

O artigo [Schram 2012] apresenta um processo de migração de um BD Relacional para um BD NoSQL Orientado a Colunas. No artigo são mostrados os benefícios adquiridos nessa migração e a importância da modelagem.

O trabalho [Schreiner 2015] propõe o SQLtoKeyNoSQL que consiste em uma camada relacional sobre BDs NoSQL para realização de operações SQL. O mapeamento do esquema relacional é feito para um modelo canônico e em seguida, é mapeado para um modelo de dados NoSQL. As diferenças entre os modelos NoSQL são abstraídas nesse modelo canônico. Assim é possível realizar operações SQL em BD NoSQL Orientados a Documentos, Orientado a Colunas e Orientado a Chave-Valor.

O artigo [Sellami 2016] propõe um modelo de dados unificador de BDs Relacionais e NoSQL (Orientado a Documentos e Chave-Valor), para que programadores possam escrever códigos de maneira independente do BD destino. Nessa proposta, as operações não são feitas diretamente no BD alvo, e sim em um armazenamento virtual de dados. No trabalho ainda é proposto um esquema hierárquico para representar a estrutura do armazenamento e também álgebra de consulta para apoio nas operações.

4. Comparação

A Tabela 1 apresenta algumas características dos artigos. Na coluna “Conceitual” é descrito o esquema conceitual utilizado no artigo, a coluna “Lógico” descreve o modelo lógico utilizado. A coluna “Físico” apresenta em quais BDs NoSQL são implementados, e para otimização do espaço, são utilizadas abreviações; Orientado a Chave-Valor; Chave-Valor, Orientado a Colunas, Colunas; Orientados a Documentos, Documentos e Orientado a Grafos, Grafos.

A coluna “Categoria” define o(s) principal(is) objetivo(s) de cada artigo: 1 para proposta de modelagem para BDs NoSQL; 2, para migração de BDs NoSQL e 3 para uma proposta de camada. Em “Categoria”, é possível encontrar mais de uma categoria

para um artigo, porém, a relevância deve ser considerada da esquerda para direita.

A figura 1, representa em forma de gráfico, o número de BDs NoSQL abordados em cada artigo. Na figura 2, o gráfico apresenta a quantidade de cada tipo de BDs NoSQL encontrados nos artigos. A figura 2 utiliza abreviações baseadas na tabela 1.

Tabela 1: Comparação entre as propostas

	Conceitual	Lógico	Físico	Categoria
[Banerjee 2015]	GOOSSDM	JSON	Documentos	1
[Bansel 2016]		Metamodelo	Colunas, Documentos e Grafos	2
[Bermbach 2015]			Colunas	1
[Bugiotti 2014]	UML	NoAM	Chave-Valor, Colunas e Documentos	1
[Feng 2015]			Colunas	1
[Fioravanti 2016]	UML		Documentos e Grafos	2, 1, 3
[Karnitis 2015]			Documentos	2
[Kaur 2013]			Documentos e Grafos	1
[Lima 2015]	EER	Modelo lógico de Documento	Documentos	1
[Schram 2012]			Colunas	2
[Schreiner 2015]		Canônico	Chave-Valor, Colunas e Documentos	3, 1
[Sellami 2016]		Modelo Unificador	Chave-Valor e Documentos	3, 1

5. Considerações finais

As aplicações atuais utilizam cada vez mais a *Cloud Computing*, que consiste no armazenamento de grandes volumes de dados em diversos servidores separados. Esse tipo de armazenamento é compatível com os BDs NoSQL, que têm como características principais a gerência de grandes volumes de dados e o apoio a escalabilidade horizontal.

Os BDs Tradicionais, os Relacionais, são estruturados e consistem no armazenamento de dados em tabelas e relacionamento entre elas. Já nos BDs NoSQL, existem quatro modelos de dados: Chave-Valor, Orientado a Colunas, Orientado a Documentos e Orientado a Grafos. Cada um dos modelos, possui uma característica própria, tornando-os compatíveis com utilizações específicas.

A eficiência e a consistência de um BD tem grande influência do seu projeto,

que consistem em 3 partes: modelo conceitual, modelo lógico e modelo físico. A modelagem lógica é onde os dados estão mais próximos para utilização, não possui uma metodologia padrão, basicamente é feita conforme as necessidades do usuário e utilizando “boas práticas”.

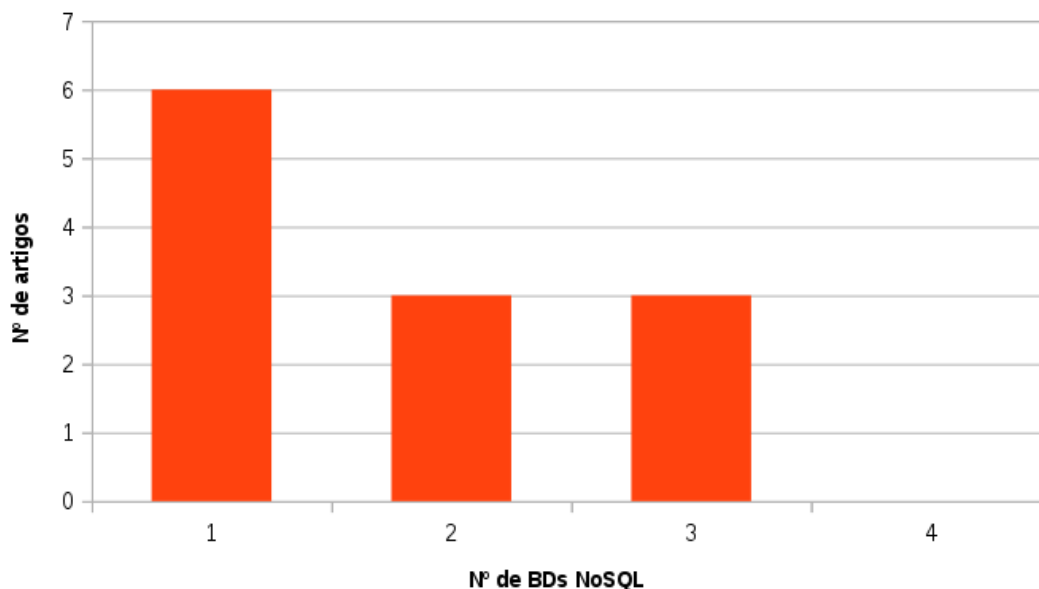


Figura 1: Número de BDs NoSQL por artigo

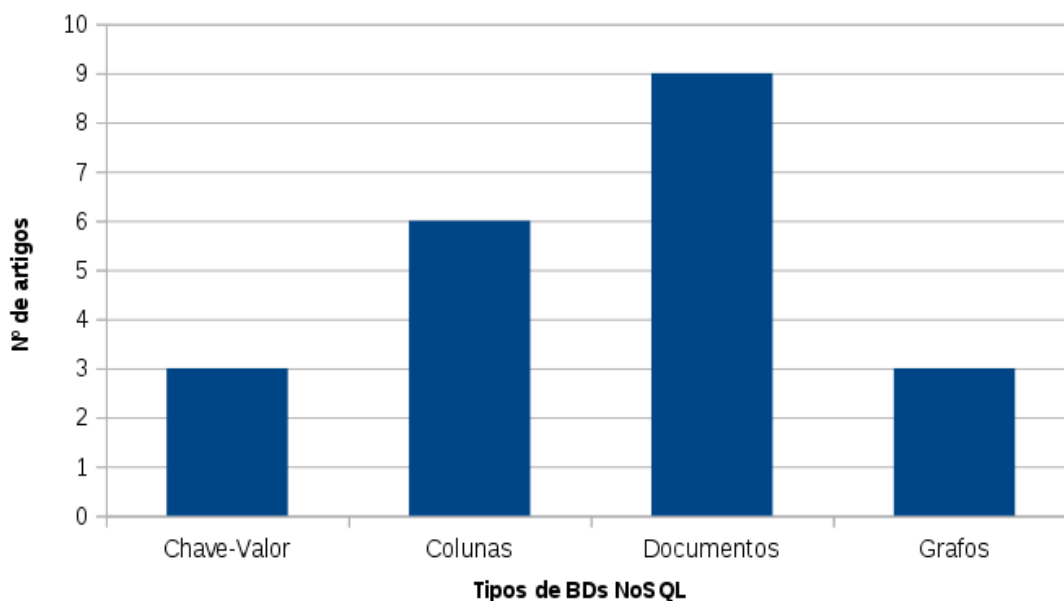


Figura 2: Quantidade de cada tipo de BD NoSQL encontrados nos artigos

A revisão busca artigos que contribuam com a modelagem lógica, seja de maneira direta ou indireta. Pois além dos artigos que são propostas de modelagem lógica para BD NoSQL, outros artigos tratam da migração de BDs ou abordam a criação de camadas entre BDs Relacionais e NoSQL, que também contribuem com a pesquisa.

Alguns trabalhos apresentam um modelo lógico que pode ser utilizado em um tipo de BD NoSQL, e outros, para alguns dos tipos. O processo de conversão do modelo

conceitual para o modelo lógico é apresentada apenas em alguns trabalhos. Pode-se concluir que não foi encontrado um trabalho que apresente uma conversão de um modelo conceitual para um modelo lógico, que possa ser utilizado em qualquer um dos quatro tipos de BDs NoSQL. Sugere-se pesquisas futuras para uma proposta de modelagem lógica para BDs NoSQL, que aborde o processo de conversão do modelo conceitual para o modelo lógico e também, que possua compatibilidade com os quatro tipos de BDs NoSQL.

Referências

- Banerjee, S., Shaw, R., Sarkar, A., and Debnath, N. C. (2015) "Towards logical level design of Big Data". In: Towards logical level design of Big Data. In 2015 IEEE 13th International Conference on Industrial Informatics (INDIN) p. 1665-1671.
- Bansel, A. and Chis, A. E. (2016). "Cloud-Based NoSQL Data Migration". In: 2016 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP) IEEE. p. 224-231.
- Bermbach, D., Eberhardt, J. and Tai, S. (2015) "Informed Schema Design for Column Store-Based Database Services". In: 2015 IEEE 8th International Conference on Service-Oriented Computing and Applications (SOCA) p. 163-172
- Bugiotti, F., Cabibbo, L. Atzeni, P. and Torlone, R. (2014). "Database design for NoSQL systems". In: International Conference on Conceptual Modeling. Springer International Publishing. p. 223-231.
- Elmasri, R. and Navathe, S. B. (2011) "Fundamentals of Database Systems". 6th edition, Pearson Addison Wesley, 2011.
- Feng, W., Gu, P., Zhang, C. and Zhou, K. (2015). "Transforming UML Class Diagram into Cassandra Data Model with Annotations". In: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). IEEE. p. 798-805.
- Fioravanti, S., Mattolini, S., Patara, F. and Vicario, E. (2016). "Experimental Performance Evaluation of different Data Models for a Reflection Software Architecture over NoSQL Persistence Layers". In: Proceedings of the 7th ACM/SPEC on International Conference on Performance Engineering. ACM. p. 297-308.
- Karnitis, G. and Arnicans, G. (2015). "Migration of Relational Database to Document-Oriented Database: Structure Denormalization and Data Transformation". In Computational Intelligence, Communication Systems and Networks (CICSyN), 2015 7th International Conference on IEEE. p. 113-118.
- Kaur, K. and Rani, R. (2013). "Modeling and querying data in NoSQL databases". In: Big Data, 2013 IEEE International Conference on IEEE. p. 1-7.
- Kitchenham, B., Brereton, P., Budgen, D., Turner, M., Bailey, J. and Linkman, S. (2009). "Systematic literature reviews in software engineering – A systematic literature review". In: Information and Software Technology, v.51, p. 7-15.
- Lima, C. and Mello, R. S. (2015). "Um Estudo sobre Modelagem Lógica para Bancos de Dados NoSQL". In: XI Escola Regional de Banco de Dados - ERBD 2015

- Lima, C. and Mello, R. S. (2015). "A workload-driven logical design approach for NoSQL document databases" . In: Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services. ACM, 2015. p. 73.
- Schram, A. and Anderson, K. M. (2012). "MySQL to NoSQL: data modeling challenges in supporting scalability". In: Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity. ACM p. 191-202.
- Schreiner, G. A., Duarte, D. and Mello, R. S. (2015). "SQLtoKeyNoSQL: a layer for relational to key-based NoSQL database mapping". In: Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services. ACM p. 74.
- Sellami, R., Bhiri, S., and Defude, B. (2016). "Supporting multi data stores applications in cloud environments". In: IEEE Transactions on Services Computing, 9(1). IEEE p. 59-71.
- Souza, A., Prado, E., Sun, V. and Fantinato, M. (2014). "Critérios para Seleção de SGBD NoSQL: o Ponto de Vista de Especialistas com base na Literatura." In Simpósio Brasileiro de Sistemas da Informação(SBSI) 2014.