

2018

Graph Databases: Contextualization and Exploratory Study

Filipe Salgado

Universidade do Minho, A69183@alunos.uminho.pt

José Luís Pereira

Universidade do Minho, jlmp@dsi.uminho.pt

Follow this and additional works at: <https://aisel.aisnet.org/capsi2018>

Recommended Citation

Salgado, Filipe and Pereira, José Luís, "Graph Databases: Contextualization and Exploratory Study" (2018). *2018 Proceedings*. 46.
<https://aisel.aisnet.org/capsi2018/46>

This material is brought to you by the Portugal (CAPSI) at AIS Electronic Library (AISEL). It has been accepted for inclusion in 2018 Proceedings by an authorized administrator of AIS Electronic Library (AISEL). For more information, please contact elibrary@aisnet.org.

Bases de Dados em Grafos: Contextualização e Estudo Exploratório.

Graph Databases: Contextualization and Exploratory Study

Filipe Salgado, Universidade do Minho, Portugal, A69183@alunos.uminho.pt

José Luís Pereira, Universidade do Minho & Centro Algoritmi, Portugal, jlmp@dsi.uminho.pt

Resumo

Depois de várias décadas de grande sucesso e bons serviços prestados às organizações, a tecnologia relacional de bases de dados tem vindo a ser desafiada por uma nova classe de tecnologias de bases de dados a que se deu a designação genérica de NoSQL (Not only SQL). Para esta situação contribuíram decisivamente os recentes desenvolvimentos na área a que se tem vindo a chamar Big Data, na qual, dada a complexidade e volume dos conjuntos de dados a gerir, o tradicional modelo relacional começou a apresentar dificuldades. Dentro desta nova classe de tecnologias de bases de dados surgiram diferentes propostas, com distintas proveniências e áreas de aplicação, vulgarmente classificadas em quatro grupos, de acordo com o seu modelo de dados: orientado a colunas, orientado a documentos, pares chave-valor e orientado a grafos. Em particular, as bases de dados em grafos apresentam um conjunto de características que permitem representar relacionamentos entre dados que nenhum outro modelo consegue fazer de forma tão capaz. Como vivemos num mundo onde a informação está toda conectada, este modelo de bases de dados tem tudo o que é necessário para ter sucesso. Sendo assim, descrevem-se alguns exemplos de aplicação desta família de base de dados assim como se demonstra a facilidade em construir *queries* que seriam extremamente complexas caso fossem realizadas através de SQL, sobre bases de dados relacionais.

Palavras-chave: *Bases de Dados em Grafos; Big Data; NoSQL; Neo4J.*

Abstract

After several decades of great success and good services to organizations, relational database technology has been challenged by a new class of database technologies usually called NoSQL (Not only SQL). The recent developments in the area called Big Data contributed decisively to this situation, in which the traditional relational model began to present difficulties, due to the complexity and large volumes of data. Within this new class of databases, different proposals, with several origins and application areas, appeared in four groups, according to their data model: column oriented, document oriented, key-value and Graphs oriented. In particular, graph databases provide a set of characteristics to represent relationships between data that no other model can represent so well. As we live in a world where information is all connected, this database model has what it takes to be successful. In this way, some examples of graph database applications will be discussed as well as demonstrations of the facility to construct queries, which would be extremely complex if they were developed in SQL over relational databases.

Keywords: *Graph Databases; Big Data; NoSQL; Neo4j.*

Agradecimento/Acknowledgement: This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013.

1. DESCRIÇÃO

Todos os dias são gerados grandes quantidades de novos dados, num fenómeno que tem vindo a ser denominado de *Big Data*. Para além do seu enorme volume estes dados são tipicamente complexos, e não estruturados. Assim sendo, a tecnologia relacional começa a apresentar grandes dificuldades no seu armazenamento e tratamento. Por essa razão, diversas outras soluções de armazenamento de dados têm vindo a surgir: as chamadas bases de dados *NoSQL*. Estas, tipicamente dividem-se em quatro grandes grupos de acordo com a forma como organizam os dados: orientadas a Colunas, orientadas a Documentos, pares Chave-Valor e orientadas a Grafos.

As bases de dados em grafos, são ainda muito pouco exploradas, contudo apresentam um potencial enorme, derivado do facto de permitirem representar perfeitamente relações complexas entre dados. Estas relações são dificilmente representadas tanto no modelo relacional como nos outros grupos de bases de dados *NoSQL*. A informação é dos recursos mais valiosos para as organizações, e se puder ser relacionada, mais valiosa se torna. Assim, pretende-se explorar as bases de dados em grafos, e verificar em que níveis estas conseguem superar o tradicional modelo relacional de bases de dados, quer em termos de capacidade de modelação, quer em termos de facilidade de utilização.

A investigação em bases de dados orientadas a grafos foi bastante popular no início da década de 1990, mas caiu em desuso por causa do *Hypertext* e da investigação em *XML* como referido por Vicknair et al. (2010). Contudo, devido ao avanço da *Web 2.0*, das redes sociais e dos dados interligados e interconectados, surgiu o interesse em desenvolver e estudar o armazenamento de dados em grafos, levando a que este tipo de base de dados fosse ganhando cada vez mais popularidade (Holzschuher & Peinl, 2013). A *Web*, na sua totalidade, é essencialmente um grafo de dados e informações ligadas em conjunto (Cudré-Mauroux & Elnikety, 2011).

Segundo Buerli & Obispo (2012), recentemente esta área tem ganho a atenção de muitos especialistas por causa de muitos projetos que estão na moda e onde é necessário armazenar, interligar e retirar valor da informação. Isto só é possível se a sua representação for feita em forma de grafos. Sendo assim, a modelação em grafos é uma alternativa viável aos modelos relacionais e são ideais para representar dados da Química, Biologia, Semântica *Web* e Redes Sociais (Miller, 2013). Segundo (Buerli & Obispo, 2012) “Nas redes sociais, os grafos são muito populares no que toca à investigação. Os grafos, não só guardam pessoas em nodos, como também fotografias, mensagens, publicações, como permitem relacionar todos estes nodos. Permitem encontrar o caminho mais perto entre dois nodos, sugerir amigos baseado nas características de um utilizador das redes sociais, e permitem também detetar padrões entre nodos ou relações”. Estas consultas tornam-se muito mais simples quando realizadas numa base de dados orientada a grafos do que num modelo relacional.

As bases de dados relacionais foram, durante muitas décadas, as escolhidas para armazenar grandes quantidades de dados devido ao seu alto desempenho e às suas propriedades *ACID*. Com o fenómeno

Big Data, novos requisitos foram emergindo, tais como os serviços em *Cloud*. Como isto, vários tipos de bases de dados *NoSQL* surgiram e estão a ganhar cada vez mais popularidade (Tudorica & Bucur, 2011). Este tipo de base de dados é útil quando se trabalha com uma enorme quantidade de dados e quando a sua natureza não requer um modelo relacional (Moniruzzaman & Hossain, 2013).

Hoje em dia, torna-se cada vez mais importante tirar proveito da análise de toda a informação que diariamente é gerada nas organizações. Como o modelo relacional está a apresentar dificuldades em armazenar e analisar estas quantidades de dados, surgiu uma nova gama de bases de dados a que se dá o nome de *NoSQL*. Estas bases de dados são, em norma, mais eficientes que as bases de dados relacionais no que toca ao armazenamento, tratamento e posterior análise de dados.

As bases de dados em grafos, destacam-se dos outros modelos *NoSQL* em termos de representação de dados interconectados, que estão cada vez mais presentes na *WEB*. Sendo assim, este modelo de base de dados apresenta um futuro promissor, ainda que neste momento não seja muito conhecido pela comunidade académica e científica. Nesse sentido, o grande objetivo deste artigo, prende-se com o facto de chamar a atenção de todos os interessados na área do *Big Data* e *NoSQL* para este modelo e para as vantagens que este traz.

O próximo passo que se encontra em desenvolvimento envolve uma comparação a nível de desempenho entre o modelo em grafos e o modelo relacional ao nível de transações CRUD (*Create, Read, Update e Delete*) com diversos tipos de testes de carga relativamente ao número de utilizadores simultâneos e diferentes volumes de dados armazenados.

REFERÊNCIAS

- Buerli, M., & Obispo, C. (2012). The Current State of Graph Databases. Department of Computer Science, Cal Poly San Luis Obispo.
- Cudré-Mauroux, P., & Elnikety, S. (2011). Graph data management systems for new application domains. *Conference on Very Large Databases*, 4(12), 1510–1511.
- Holzschuher, F., & Peinl, R. (2013). Performance of graph query languages: comparison of Cypher, Gremlin and native access in Neo4j. *16th International Conference on Extending Database Technology, EDBT' 13*, (March 2013), 195–204.
- Moniruzzaman, A. B. M., & Hossain, S. A. (2013). Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. *Nosql Database: New Era of Databases for Big Data Analytics-Classification, Characteristics and Comparison*, 6(4), 1–14.
- Vicknair, C., Macias, M., Zhao, Z., Nan, X., Chen, Y., & Wilkins, D. (2010). A comparison of a graph database and a relational database. *Proceedings of the 48th Annual Southeast Regional Conference on ACM SE 10*, 1.