

Oportunidades de Pesquisa para o Uso de Infraestruturas Blockchain na Gestão de Dados Distribuídos

Leonardo O. Moreira^{1,2}, Carlos S. S. Marinho^{2,3}, Maurício Moreira Neto^{1,3}, Emanuel F. Coutinho¹, José Neuman de Souza³, Javam C. Machado^{2,3}

¹Grupo de Pesquisa em Computação em Nuvem e Sistemas (IBITURUNA)
Universidade Federal do Ceará (UFC) – Fortaleza, CE – Brasil

²Laboratório de Sistemas e Banco de Dados (LSBD)
Universidade Federal do Ceará (UFC) – Fortaleza, CE – Brasil

³Mestrado e Doutorado em Ciência da Computação (MDCC)
Universidade Federal do Ceará (UFC) – Fortaleza, CE – Brasil

{leoomoreira, sergio.marinho}@lsbd.ufc.br, maumneto@alu.ufc.br

{emanuel.coutinho, neuman}@ufc.br, javam.machado@lsbd.ufc.br

Abstract. *A wide variety of applications are data driven. Thus, several concerns and care with the management of these data should be taken into account. Some applications may require different data requirements, such as availability, integrity, transparency, audibility, immutability, and irrefutability. Such requirements can be found in Blockchain infrastructures. Then, when adopting Blockchain in the solutions of these applications, these characteristics are guaranteed by the characteristics of the Blockchain infrastructures. This article shows some research opportunities involving distributed data management in Blockchain.*

Resumo. *Uma grande variedade de aplicações é orientada a dados. Assim, diversas preocupações e cuidados com a gestão destes dados devem ser levadas em consideração. Algumas aplicações podem necessitar de diferentes requisitos de dados, tais como disponibilidade, integridade, transparência, auditabilidade, imutabilidade e irrefutabilidade. Tais requisitos podem ser encontrados em infraestruturas de Blockchain. Então, ao adotarem a Blockchain nas soluções destas aplicações, essas características são garantidas pela infraestrutura da Blockchain. Este artigo mostra algumas oportunidades de pesquisas envolvendo a gestão de dados distribuídos em Blockchain.*

1. Introdução

Como muitas aplicações são orientadas a dados [Maenhaut et al. 2015], sistemas de gestão de dados surgem como soluções para aprimorar a manipulação de dados em tais aplicações. Diante do crescimento do poder computacional e da conectividade, um grande volume de dados está sendo gerenciado por aplicações e possuem as mais diversas características [Hashem et al. 2015]. Para gerenciar dados em larga escala surge a necessidade da adoção de técnicas de distribuição dos dados para aumentar a disponibilidade, desempenho [Rabl and Jacobsen 2017], confiabilidade e segurança [Alsirhani et al. 2017]. Além disso, algumas soluções para gestão de dados devem considerar que certas aplicações possuem requisitos de dados diferentes.

Blockchain é uma tecnologia descentralizada e imutável, em que os dados são armazenados permanentemente e sequencialmente por transações distribuídas [Chakraborty et al. 2018]. Ademais, aspectos de replicação de dados entre os participantes da rede estão presentes em Blockchain [Nakamoto 2008], o que aumenta a disponibilidade dos dados e proporciona auditabilidade [Greve et al. 2018]. Além disso, por conta do uso de aspectos de criptografia, na Blockchain, um nó não pode negar que enviou uma transação realizada por ele mesmo [Braga et al. 2017]. Essas características podem ser requisitos importantes para algumas classes de aplicações orientadas a dados.

O objetivo deste artigo é apresentar pesquisas que fazem uso de infraestruturas Blockchain para garantir, nas aplicações, algumas características importantes na gestão de dados distribuídos. Para isso, a tecnologia Blockchain e alguns aspectos relacionados à gestão de dados distribuídos serão apresentados. Em seguida, algumas oportunidades de pesquisa envolvendo o uso de Blockchain na gestão de dados distribuídos serão identificadas. Por fim, serão expostas as conclusões da presente pesquisa.

2. Blockchain e Gestão de Dados Distribuídos

Blockchain é uma tecnologia que fornece suporte distribuído, confiável e seguro para a realização de transações entre participantes dispersos em uma rede *Peer-to-Peer* (P2P) de larga escala. Na rede Blockchain, os nós não obrigatoriamente possuem confiança entre si e não há a existência de uma terceira parte confiável, como cartórios, para intermediar as relações entre os nós [Greve et al. 2018, Yaga and Mell 2018]. Para propor uma nova transação, a partir de um programa cliente, o usuário deve se comunicar com algum nó que participe da rede, informando-o sobre a transação proposta. Em seguida, o nó repassa a transação para os demais nós da rede, e então, é decidido, por meio de um algoritmo de consenso, a ordem de armazenamento das transações. As transações são armazenadas em blocos, que são encadeados, formando uma estrutura denominada Blockchain. Em cada nó da rede, há uma réplica dessa cadeia de blocos em uma estrutura denominado *ledger* [Nakamoto 2008].

As Blockchains possuem propriedades que contribuem para o desenvolvimento de aplicações. Uma delas é a descentralização, que ocorre em virtude de sistemas serem executados de forma distribuída, sem a existência de uma terceira parte confiável. A disponibilidade é outra característica, pois os dados e transações são replicados em todos os nós da rede, mantendo o sistema mais disponível. Blockchains também fornecem transparência e auditabilidade, visto que as transações registradas no *ledger* são visíveis para todos os nós da rede, de modo que possam ser inspecionados. A imutabilidade é a propriedade que garante que não deve ser possível mudar uma transação após registrá-la na rede. Por último, a irrefutabilidade assegura que um nó não pode negar que enviou uma transação criada por ele mesmo [Chowdhury et al. 2018, Greve et al. 2018, Braga et al. 2017, Zheng et al. 2017].

3. Oportunidades de Pesquisa

Uma possibilidade a ser explorada é o desenvolvimento de abordagens para gerenciar dados, de forma híbrida e transparente, para as aplicações: que utilize o modelo Relacional e Blockchain. O propósito é que a aplicação se comunique via SQL com a abordagem híbrida de gerenciamento de dados, de modo a fornecer transparência para

a aplicação. Assim, é possível que uma aplicação use um conjunto de dados mais adequado ao modelo Relacional, e um outro mais pertinente ao contexto Blockchain. Dados mais característicos de Blockchain estão em contextos em que não há uma terceira parte confiável, além de não haver confiança entre os nós da rede ou imutabilidade dos dados. Por outro lado, dados que são mais adequados ao modelo Relacional são aqueles os quais se deseja realizar consultas mais complexas, com o uso de junções ou aspectos analíticos.

Infraestruturas de Blockchain não são projetadas para armazenar dados relacionais e deve haver alguma mudança na estrutura do dado recebido, mantendo sua função semântica, para que ele possa ser armazenado em uma Blockchain. Além disso, geralmente, em Blockchain, um dado é salvo e possui um identificador *hash* para ser encontrado. Já em Banco de Dados Relacional, são utilizadas outras formas para identificar dados, como números e palavras, por exemplo. Assim, é necessário desenvolver algum mecanismo para se fazer correspondência do identificador escolhido pelo usuário, via SQL, com o identificador em uso na Blockchain. Se esse mapeamento não for feito, é provável que seja necessário verificar todos os registros de dados da Blockchain para encontrar um arquivo único desejado.

A tecnologia Blockchain possui características vantajosas em aplicações que utilizam dados sensíveis, como aplicações de *E-health*. As aplicações *E-health* são soluções computacionais para problemas relacionados às áreas da saúde, desde o monitoramento de dados fisiológicos de um usuário até a gestão de registros de pacientes em uma infraestrutura hospitalar. Dados fisiológicos e laudos médicos devem ser acessados somente por usuários seletos, como os profissionais da saúde. Assim, Blockchains podem fornecer características, como segurança e imutabilidade, aos dados de saúde, de modo que não seja possível excluir ou modificar uma informação que já tenha sido validada pela rede. Uma nova informação pode ser adicionada em um próximo bloco como uma informação adicional ao histórico do registro. Uma linha de pesquisa importante nesse contexto é a privacidade dos dados em uma Blockchain. Os dados médicos, como o laudo de um paciente, suas comorbidades e a sua sequência de DNA, são exemplos de dados sensíveis que devem ser privativos e acessados somente por usuários autorizados. Porém, em uma Blockchain pública, qualquer usuário pode ter acesso aos dados. Logo, o estudo da utilização de técnicas, aplicadas a dados sensíveis, como anonimização, busca criptográfica, *Private Information Retrieval* (PIR) [Branco Junior et al. 2014] e outras em uma Blockchain se torna um ponto relevante para pesquisas em *E-health*.

4. Conclusão

Este trabalho apresentou algumas oportunidades de pesquisa que fazem o uso da infraestrutura Blockchain para prover características de gestão de dados distribuídos nas aplicações. Para isso, foram apresentados, de modo sucinto, a tecnologia Blockchain e aspectos da gestão de dados distribuídos de um modo geral. Na sequência, foram discutidas oportunidades de pesquisa envolvendo o uso de Blockchain na gestão de dados distribuídos. É possível perceber que a tecnologia Blockchain fornece características que facilitam o desenvolvimento, como: disponibilidade, confiabilidade, auditabilidade, irrefutabilidade e replicação. Isso diminui o esforço, por parte dos desenvolvedores, no projeto e implementação de aplicações orientadas a dados que requerem as características presentes na Blockchain. Com isso, emerge uma gama de outras oportunidades de pesquisa com Blockchain e gestão de dados.

Referências

- Alsirhani, A., Bodorik, P., and Sampalli, S. (2017). Improving database security in cloud computing by fragmentation of data. In *2017 International Conference on Computer and Applications (ICCA)*, pages 43–49. IEEE.
- Braga, A. M., Marino, F. C. H., and dos Santos, R. R. (2017). *Segurança de Aplicações Blockchain Além das Criptomoedas*, chapter 3, pages 100–148. SBC.
- Branco Junior, E. C., Machado, J. C., and Monteiro, J. M. (2014). *Estratégias para Proteção da Privacidade de Dados Armazenados na Nuvem*, chapter 2, pages 46–75. SBC.
- Chakraborty, P., Shahriyar, R., Iqbal, A., and Bosu, A. (2018). Understanding the software development practices of blockchain projects: A survey. In *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM '18*, pages 28:1–28:10, New York, NY, USA. ACM.
- Chowdhury, M. J. M., Colman, A., Kabir, M. A., Han, J., and Sarda, P. (2018). Blockchain versus database: A critical analysis. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 1348–1353. IEEE.
- Greve, F., Sampaio, L., Abijaude, J., Coutinho, A., Ítalo Valcy, and Queiroz, S. (2018). *Blockchain e a Revolução do Consenso sob Demanda*, chapter 5, pages 1–52. SBC.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information systems*, 47:98–115.
- Maenhaut, P., Moens, H., Ongenae, V., and De Turck, F. (2015). Design and evaluation of a hierarchical multi-tenant data management framework for cloud applications. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pages 1208–1213.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
- Rabl, T. and Jacobsen, H.-A. (2017). Query centric partitioning and allocation for partially replicated database systems. In *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD '17*, pages 315–330, New York, NY, USA. ACM.
- Yaga, D. and Mell, P. (2018). *NISTIR 8202: Blockchain Technology Overview*, chapter 1, pages 1–68. National Institute of Standards and Technology.
- Zheng, Z., Xie, S., Dai, H., Chen, X., and Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. In *2017 IEEE International Congress on Big Data (BigData Congress)*, pages 557–564.