

Um Estudo Exploratório entre Banco de Dados NoSQL e Armazenamento de Dados em Blockchain

José Cleanto F. Arrais Neto¹, Francisca Luzia N. Araújo²,
Maurício Moreira Neto^{3,4}, Gabriel A. L. Paillard^{1,2}, Leonardo O. Moreira¹

¹Instituto Universidade Virtual (IUVI)
Universidade Federal do Ceará (UFC) – Fortaleza, CE – Brasil

²Programa de Pós-Graduação em Computação (PCOMP)
Universidade Federal do Ceará (UFC) – Quixadá, CE – Brasil

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

⁴Centro Universitário Christus (UNICHRISTUS) – Fortaleza, CE – Brasil

cleantoneto2010@hotmail.com, luzia.nog94@gmail.com,
maumneto@alu.ufc.br, {gabriel, leomoreira}@virtual.ufc.br

Abstract. *Several data models are being proposed, because in some application classes the relational model is not so efficient. Some of these proposed data models can be found and implemented in NoSQL Databases. In addition, other technologies are emerging, such as Blockchain, which store data aiming at other types of characteristics that are important for applications. This paper aims to present an initial exploratory study to support a future deeper comparative analysis between NoSQL Database and Blockchain data storage. To this end, we bring an analysis of the characteristics of NoSQL and Blockchain with the data requirements of two applications.*

Resumo. *Vários modelos de dados estão sendo propostos, pois em algumas classes de aplicações o modelo relacional não se mostra tão eficaz. Alguns destes modelos de dados propostos podem ser encontrados e implementados em Bancos de Dados NoSQL. Além disso, estão surgindo outras tecnologias, como Blockchain, que fazem armazenamento de dados visando outros tipos de características importantes para as aplicações. Este artigo tem como objetivo apresentar um estudo exploratório inicial para suportar uma futura análise comparativa mais aprofundada entre Banco de Dados NoSQL e armazenamento de dados em Blockchain. Para este fim, trouxemos uma análise sobre as características de NoSQL e Blockchain com os requisitos de dados de duas aplicações.*

1. Introdução

Um Banco de Dados pode ser definido como uma coleção de dados relacionados, onde o modelo relacional é um modelo de dados amplamente utilizado por aplicações de diversos domínios [Elmasri e Navathe 2011]. Este modelo de dados, o relacional, estrutura um Banco de Dados como uma coleção de uma ou mais relações (tabelas) compostas

de tuplas (linhas) e atributos (colunas) [Marinho et al. 2020]. Como muitas aplicações são orientada a dados, um grande volume de dados é produzido e se faz necessário um processamento adequado, visando alguns aspectos como desempenho, disponibilidade e segurança. Assim, outros modelos de dados têm sido propostos visando algumas classes de aplicações, onde o modelo de dados relacional não se mostra eficaz. Outro fator que deve-se levar em consideração, é que o maior tempo gasto no processamento das requisições de aplicações está relacionada à solução de persistência de dados [Moreira 2014]. Portanto, em aplicações orientadas a dados, deve-se ter uma maior atenção com o modelo de dados utilizado e também com o componente que implementa a solução de persistência de dados.

Marinho et al. (2020) comentam que muitas aplicações, orientadas a dados, utilizam dados de diferentes tipos e domínios. Assim, alternativas ao modelo de dados relacional, como o *Not Only SQL* (NoSQL), estão se solidificando [Marinho et al. 2020]. MongoDB [MongoDB 2021b] e Blockchain [Moreira Neto et al. 2020] são, respectivamente, um Banco de Dados NoSQL distribuído de uso geral baseado em documentos e uma tecnologia que permite armazenamento de dados distribuídos. O MongoDB, além de ser um Banco de Dados NoSQL popular, é frequentemente apreciado em estudos comparativos devido à sua flexibilidade e facilidade de uso [Andreoli et al. 2021].

El-Hindi et al. (2019) discutem que o BigchainDB [BigchainDB 2021a] se baseia no MongoDB, assim possuindo características semelhantes. BigchainDB, além das características comuns de um Banco de Dados, acrescenta os atributos vinculados à tecnologia Blockchain. Portanto, estas duas soluções de persistência de dados, MongoDB e BigchainDB, possuem algumas similaridades, mas com propostas de gestão de dados diferentes. Diante disso, a principal contribuição deste artigo é confrontar as duas soluções supracitadas de persistência de dados e discutir quais benefícios podem trazer às aplicações orientadas a dados.

Segundo Gil (2002), algumas regras práticas para formulação de problemas científicos são: i) deve ser estruturada como uma pergunta; ii) deve ser a mais específica possível; e iii) utilizar terminologias claras com significativo preciso. Assim, pode-se formular o problema científico que permeia este trabalho com a seguinte questão: “Como um Banco de Dados NoSQL e o armazenamento de dados em Blockchain podem beneficiar aplicações orientadas a dados?”. Além da questão que reflete o problema científico desta pesquisa, são formuladas duas hipóteses: i) as duas soluções de persistência de dados possuem características diferentes quanto a gestão de dados; e ii) cada uma das duas soluções de persistência podem apresentar vantagens e desvantagens dependendo dos requisitos e da natureza das aplicações orientadas a dados.

O objetivo principal deste ensaio é apresentar um estudo exploratório inicial, na perspectiva da gestão de dados, para suportar futuros aprofundamentos de análises comparativas entre Banco de Dados NoSQL e Blockchain, visando benefícios às aplicações. Para alcançar o objetivo principal supracitado, os seguintes objetivos secundários foram elencados: i) apresentar os conceitos e motivações relacionadas às tecnologias de Banco de Dados NoSQL e Blockchain; ii) elaborar uma metodologia de natureza exploratória no intuito de confrontar as duas soluções de persistência de dados; e iii) realizar uma discussão comparativa, entre as duas soluções de persistência de dados, diante do estudo exploratório conduzido e dos requisitos de dados das aplicações.

Este trabalho está dividido, além da introdução, nas seguintes seções: Seção 2 apresenta todo arcabouço teórico necessário para o entendimento, compreendendo os conceitos de Banco de Dados Relacional, Banco de Dados NoSQL e Blockchain. Ainda na Seção 2 são descritos o MongoDB e BigchainDB como soluções, respectivamente, de Banco de Dados NoSQL e Banco de Dados com Características de Blockchain. Já a Seção 3 descreve os aspectos metodológicos adotados no trabalho. A Seção 4, por sua vez, comenta sobre os resultados obtidos e apresenta a discussão. Por fim, a Seção 5 apresenta as conclusões e comenta sobre trabalhos futuros.

2. Fundamentação Teórica

Nesta seção foi feita uma exploração teórica por meio de um levantamento bibliográfico sobre comparações entre Banco de Dados Relacional e Banco de Dados NoSQL, o Banco de Dados NoSQL MongoDB, a tecnologia Blockchain na perspectiva do armazenamento de dados e o Banco de Dados em Blockchain BigchainDB. Todo o arcabouço teórico escrito nesta seção é de suma importância para melhor compreensão dos resultados alcançados e a discussão, elementos textuais presentes na Seção 4.

2.1. Banco de Dados Relacional e Banco de Dados NoSQL

Com a popularização da Internet e do surgimento de infraestruturas como a Computação em Nuvem, surge a necessidade de utilizar Bancos de Dados que possam armazenar e processar Big Data de forma eficaz, com demanda por alto desempenho na leitura e escrita dos dados [Han et al. 2011]. Big Data, de forma resumida, pode ser utilizado para coletar, tratar, analisar e processar grandes volumes de dados, de diversos tipos, em altíssima velocidade. Assim, tendo em vista o crescimento e a variedade de tipos de dados gerados a todo instante, o modelo de dados relacional nem sempre é o mais recomendado, principalmente, quando grandes volumes de dados devem ser gerenciados [Soares e Boscaroli 2013]. Portanto, um Banco de Dados Relacional enfrenta muitos desafios [Han et al. 2011] diante das situações supracitadas.

Um Banco de Dados que implementa o modelo de dados relacional pode ser definido como um conjunto de relações (tabelas), colunas (atributos), tuplas (linhas) e as restrições de integridade do modelo relacional [Elmasri e Navathe 2011]. O estudo conduzido por Farias (2014) destaca algumas limitações do modelo de dados relacional: i) não permite uma estrutura de dados flexível que favoreça a obtenção de um nível mais alto de escalabilidade; e ii) não possibilita a estruturação dos seus dados de acordo com as necessidades individuais. Apesar do Banco de Dados Relacional ainda ser um dos mais utilizados, ele possui algumas limitações que podem impactar negativamente em certas aplicações que necessitem uma estrutura de dados flexível e processamento em larga escala.

Han et al. (2011) discutem que as principais vantagens dos Bancos de Dados que implementam NoSQL são as seguintes: i) leitura e escrita de dados eficiente; ii) suporte ao armazenamento em massa; iii) fácil de expandir; e iv) baixo custo. Vale ressaltar que os Bancos de Dados NoSQL também apresentam algumas desvantagens: i) não suportam a linguagem *Structured Query Language* (SQL) que é padrão da indústria; ii) ausência de transações, relatórios e outros recursos adicionais; e iii) não é um modelo de dados consolidado o suficiente para a maioria dos produtos [Han et al. 2011].

Além das desvantagens do NoSQL supracitadas, Sadalage e Fowler (2013) discutem quais transações em Banco de Dados Relacional permitem a manipulação de qualquer combinação de linhas e quaisquer tabelas em uma única transação. Tais transação implementam as propriedades chamadas de Atomicidade, Consistência, Isolamento e Durabilidade (ACID). ACID é um acrônimo, onde propriedades como a atomicidade significa que muitas linhas por muitas tabelas são atualizadas em uma única operação. Ou a operação tem sucesso total, ou desfaz totalmente os efeitos produzidos. Operações concorrentes ficam isoladas umas das outras em um ambiente de múltiplos usuários, de modo que não possam ver uma atualização parcial [Sadalage e Fowler 2013]. Note que as transações ACID aumentam a integridade e confiabilidade dos dados, alguns pontos fortes dos Banco de Dados Relacional.

Em busca por soluções para problemas pontuais, as implementações de Banco de Dados NoSQL acabam se distanciando umas das outras, o que gera diferentes estruturas [Freitas et al. 2015]. As estruturas são associadas aos modelos de dados, que no NoSQL são categorizados em chave-valor, colunas, documentos e grafos. Freitas et al. (2015) fazem um síntese sobre estes modelos:

- **Chave-valor.** É o modelo que possui a estrutura mais simples, onde sua representação estrutural é constituída de uma lista de pares de itens compostos por uma chave e um valor [Freitas et al. 2015]. O modelo chave-valor pode ser comparado à estrutura de dados Tabela Hash, onde os valores são indexados às chaves de busca e permitem um rápido acesso aos seus valores. Apesar de ser o mais simples, o modelo chave-valor é o que suporta maior volume de dados e potenciais para escalabilidade [Farias 2014]. No entanto, o modelo chave-valor não permite que consultas sejam realizadas sobre os dados, todo acesso é feito por meio de chaves de busca. Além disso, o modelo chave-valor não agrupa dados por instâncias, todos os dados estão armazenados em uma simples e não existe a possibilidade de formular consultas mais complexas, como subconsultas [Freitas et al. 2015].
- **Colunas.** É o modelo que mais se assemelha ao modelo relacional, pois sua organização é baseada em linhas e colunas [Freitas et al. 2015]. Segundo Freitas et al. (2015) esse modelo de dados não trata os dados de forma normalizada e, geralmente, não visa a consistência das informações. O modelo em colunas facilita a execução de consultas em um subconjunto de dados, mas não permite consultas onde se faz necessário a junção de famílias de colunas, pois este modelo não faz uso de recursos de chaves estrangeiras [Freitas et al. 2015].
- **Documentos.** Da mesma forma que o modelo chave-valor, o modelo baseado em documentos também faz uso de associações entre pares de chaves e valores, mas os dados não estão dispostos em uma única estrutura de dados. No modelo de documentos os dados estão agrupados em documentos que podem seguir as regras estruturais do *eXtensible Markup Language (XML)* e *JavaScript Object Notation (JSON)*, onde a última é a mais adotada [Freitas et al. 2015]. Um documento pode ser caracterizado como um conjunto de chaves e valores que estão relacionados a uma instância de dados. Assim, vários documentos pertencentes a um mesmo domínio são armazenados em uma coleção de documentos, da mesma forma como as tuplas (linhas) são armazenadas em tabelas (relações) no modelo relacional [Freitas et al. 2015]. Diferentemente dos demais modelos NoSQL, o

modelo baseado em documentos suporta referências e garante a integridade referencial [Freitas et al. 2015].

- **Grafos.** Em comparação aos modelos NoSQL anteriores, o modelo em grafos é o que mais difere dos demais. Os modelos de dados anteriores visam estratégias de armazenamento de dados, o modelo em grafos se destaca pelos relacionamentos que ocorrem entre entidades de sua base [Freitas et al. 2015]. O modelo em grafo toma como base a teoria dos grafos, onde segue o uso de três tipos de elementos: nós, arestas e propriedades. Os nós representam as instâncias, as arestas correspondem aos relacionamentos entre instâncias e as propriedades dizem respeito aos valores de dados contidos nas instâncias, os quais podem ser, por exemplo, lógicos, inteiros, caracteres e conjuntos de valores [Freitas et al. 2015]. Freitas et al. (2015) discutem quão diferente dos demais modelos, o modelo em grafos suporta o conceito de referências, garantindo a integridade referencial, o que assegura que o nó de entrada sempre faça referência ao nó de saída.

2.2. MongoDB

O MongoDB¹ é um popular Banco de Dados NoSQL de código-aberto, que implementa o modelo de dados orientado a documentos e frequentemente apreciado em estudos comparativos devido à sua flexibilidade e facilidade de uso, ao mesmo tempo que oferece todos os recursos necessários para atender aos requisitos complexos de aplicativos modernos [Andreoli et al. 2021]. Andreoli et al. (2021) destacam que MongoDB possui uma capacidade de armazenar e servir grandes volumes de dados. Assim, por ser capaz de lidar com tráfego de grandes volumes de dados com eficiência, destaca-se de Banco de Dados Relacionais que não são capazes de alcançar facilmente esse tratamento eficaz. Inclusive no trabalho conduzido por Andreoli et al. (2021) é comentado, por meio do estudo conduzido por Rubio et al. (2020), que a velocidade do MongoDB é 10 vezes maior do que a do Banco de Dados Relacional MySQL, quando o volume de dados ultrapassa 50 GB.

O MongoDB faz a gestão de documentos em um formato chamado *binary JSON* (BSON), uma estrutura de serialização codificada em JSON otimizado para velocidade, espaço e flexibilidade. Possui uma linguagem de consulta rica e expressiva que permite filtrar e classificar por qualquer campo, não importa o quão aninhado ele possa estar em um documento. As próprias consultas são JSON e, portanto, facilmente combináveis [MongoDB 2021b]. É válido ressaltar que o MongoDB armazena os dados no formato BSON internamente e pela rede, mas isso não significa que não se possa pensar no MongoDB como um Banco de Dados JSON. Qualquer coisa que você possa representar em JSON pode ser armazenado nativamente no MongoDB e recuperado com a mesma facilidade em JSON [MongoDB 2021a].

Nas versões mais recentes do MongoDB, adicionou-se o suporte a transações ACID distribuídas de vários documentos com *Snapshot Isolation*, onde por definição as transações fornecerão uma visão globalmente consistente dos dados e forçarão a execução de tudo ou nada para manter a integridade dos dados [MongoDB 2021b]. O MongoDB suporta redundância e alta disponibilidade por meio da implantação de um conjunto de réplicas e um grupo de instâncias que mantém o mesmo conjunto de dados. Em resumo, o MongoDB permite que os desenvolvedores de aplicações orientadas a dados implementem com facilidade, adaptem-se rapidamente e escalem de forma confiável.

¹MongoDB. Disponível em: <https://www.mongodb.com/>. Acessado em: 10 de junho de 2021.

2.3. Armazenamento de Dados em Blockchain

A tecnologia Blockchain fornece suporte a transações distribuídas, confiáveis e seguras aos participantes de um rede *peer-to-peer* (P2P) em larga escala [Moreira Neto et al. 2020]. Uma transação é o registro de informações de algum ativo na rede. Um ativo pode representar qualquer objeto físico ou digital que contém dados que são imutáveis. Este registro é composto pelo ativo, junto com dados sobre quem enviou, quem recebeu, horário da transação e dentre outros. A Figura 1 ilustra um exemplo de arquitetura de redes P2P com a Blockchain. Cada nó computacional possui o livro-razão, comumente conhecido como *ledger*, que armazenada todas as transações que ocorreram na Blockchain e todos os nós se comunicam entre si por meio desta arquitetura.

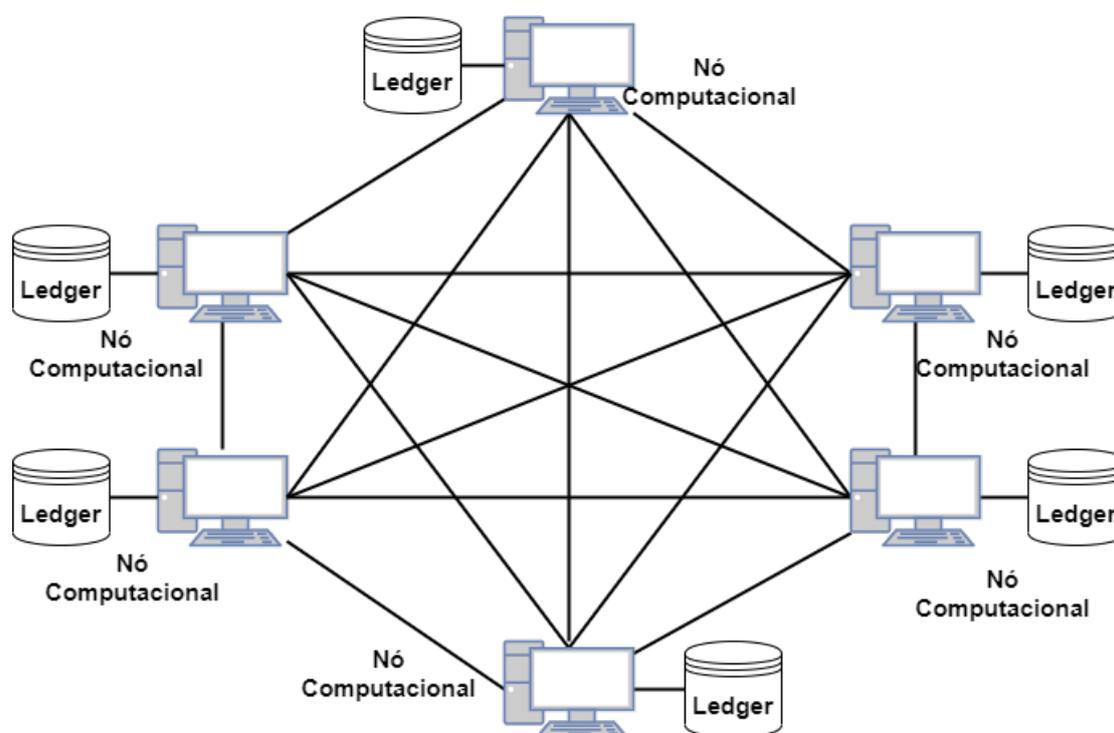


Figura 1. Arquitetura de Redes P2P da Blockchain (Adaptado de [Braga et al. 2017])

A Blockchain é uma tecnologia disruptiva, pois os nós computacionais da rede não confiam necessariamente uns nos outros e não se tem um elemento central para intermediar as transações que ocorrem na rede [Chen et al. 2018]. A Blockchain é um sistema descentralizado e com alta disponibilidade, pois estará disponível mesmo se alguns nós computacionais estejam desconectados da rede [Yaga e Mell 2018]. Uma das características da Blockchain é a capacidade de auditar as transações realizadas na rede [Chen et al. 2018]. Todas as transações são registradas no livro-razão, e por serem visíveis a todos os membros da rede, é possível inspecioná-los [Yaga e Mell 2018]. Além disso, as transações da Blockchain são imutáveis, ou seja, uma vez que a transação é validada pela rede e adicionada ao *ledger*, não pode ser alterada [Moreira Neto et al. 2020]. A Figura 2 apresenta um exemplo estrutural de uma Blockchain genérica. Cada bloco da Blockchain contém um conjunto de transações e um ponteiro (*hash*) que aponta para o bloco anterior. Esta estrutura de dados é replicada em todos os nós da rede.

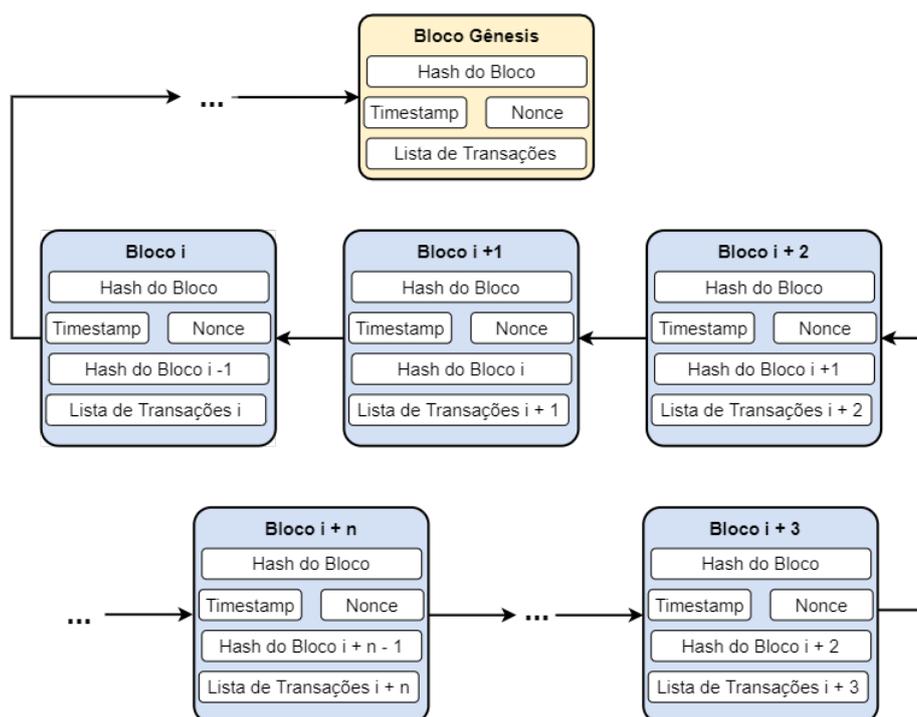


Figura 2. Exemplo de uma Estrutura Blockchain (Adaptado de [Nofer et al. 2017])

De uma forma resumida, seguem as características de uma Blockchain:

- **Disponibilidade e Integridade.** Os nós possuem os dados replicados do *ledger*, o que garante a integridade e a consistência dos dados armazenados na Blockchain;
- **Transparência e Auditabilidade.** Todas as informações contidas no *ledger* são públicas e podem ser analisadas e auditadas;
- **Imutabilidade e Irrefutabilidade.** As informações registradas no *ledger* não podem ser alteradas e as atualizações somente são possíveis a partir da geração de novas transações. O usuário ou cliente da Blockchain, ao inserir uma transação na Blockchain, não pode informar que não foi o responsável pela transação, implementando a característica de irrefutabilidade;
- **Privacidade e Anonimidade.** Greve et al. (2018) enfatizam que cada nó armazena uma parte criptografada de dados dos usuários. A anonimização dos dados transacionados são parciais, já que a visibilidade da transação é pública;
- **Desintermediação.** Greve et al. (2018) comentam que a Blockchain possibilita a integração entre diversos sistemas de forma direta e eficiente;
- **Cooperação e Incentivos.** O modelo de negócios de uma Blockchain é baseada em incentivos [Greve et al. 2018]. Quando um nó consegue validar um bloco, uma parte do ativo é fornecido a este nó.

A Blockchain utiliza algoritmos de consenso para validar a inserção de um novo bloco [Fan et al. 2020]. Existem diversos algoritmos de consenso, sendo os mais comuns o Prova de Trabalho (Do inglês, *Proof of Work*, PoW) e a Prova de Aposto (Do inglês, *Proof of Stake*, PoS) [Yaga e Mell 2018]. A PoW é um mecanismo de consenso descentralizado que exige que os nós computacionais que constituem a rede se esforcem para resolver um problema matemático arbitrário [Yang et al. 2019]. Entretanto, o mecanismo

de PoW é computacionalmente oneroso e exige um gasto de energia muito alto dos nós para minerar um único bloco da rede. O mecanismo de PoS surgiu como uma alternativa a PoW. O mecanismo de PoS seleciona os validadores por meio da proporção da quantidade de participações deste nó na Blockchain [Niya et al. 2019]. A PoS tem um gasto de energia menor do que o mecanismo PoW [Niya et al. 2019].

As Blockchains podem ser divididas em dois tipos de acordo com o nível de acesso: Blockchain pública e Blockchain privada. A Blockchain pública, também conhecida como Blockchain *permissionless*, é caracterizada pela ausência de permissão e com acesso aberto. Qualquer usuário pode entrar e participar de uma Blockchain pública, além de permitir a leitura e gravação de transações [Aleksieva et al. 2020]. Em contrapartida, as Blockchains privadas, também conhecida como Blockchain *permissioned*, impedem que usuários não autorizados acessem a rede [Behl et al. 2020] [Aleksieva et al. 2020]. Estas Blockchains utilizam controles de acesso por meio de uma entidade central, que é um nó computacional responsável por autorizar o acesso aos usuários, ou seja, apenas usuários que são autorizados a compor a rede podem ler e gravar transações [Aleksieva et al. 2020].

Os contratos inteligentes, presentes em Blockchains, são programas, escritos em linguagens de programação, implantados e executados em cada um dos nós computacionais da rede Blockchain [Braga et al. 2017]. Segundo Braga et al. (2017), um contrato inteligente consiste em programas seguros e imparáveis que representam acordos executáveis e exigíveis automaticamente. Os contratos inteligentes resolvem questões ou desafios que necessitam de acordos (consensos) servindo-se de uma mínima confiança entre os nós participantes de sistemas distribuídos [Braga et al. 2017]. Braga et al. (2017) comentam que o programa executável (binário), que representa o contrato inteligente, é implantado e executado por todos os nós computacionais da rede P2P de uma Blockchain e sua execução correta é garantida pelos mecanismos de consenso.

2.4. BigchainDB

O BigchainDB² é um Banco de Dados Distribuído, que implementa os conceitos de Big Data e adiciona características de Blockchain, possuindo um controle descentralizado, imutabilidade e transferência de ativos digitais [BigchainDB 2021b]. Um ativo pode representar qualquer objeto físico ou digital que contém dados que são imutáveis. Segundo El-Hindi (2019), o BigchainDB se baseia no Banco de Dados NoSQL MongoDB, o que pode ajudar na análise comparativa entre estas duas soluções. O BigchainDB é uma solução que pode ser utilizada por desenvolvedores e organizações que procuram um Banco de Dados consultável com características de Blockchain e a capacidade de tratar qualquer coisa armazenada no Banco de Dados como um ativo. Quer sejam átomos, *bits* ou *bytes* de valor [BigchainDB 2021b]. BigchainDB (2021b) sumariza algumas características do BigchainDB:

- **Descentralização.** Não existe um ponto único de controle, assim nenhum ponto único de falha. O controle descentralizado por meio de uma federação de nós de votação contribui para uma rede P2P;
- **Imutabilidade.** É mais do que apenas resistente à violação. Uma vez armazenados, os dados não podem ser alterados ou excluídos;

²BigchainDB. Disponível em: <https://www.bigchaindb.com/>. Acessado em: 10 de junho de 2021.

- **Tolerante a Falhas Bizantinas.** Até um terço dos nós na rede pode estar experimentando falhas arbitrárias e o resto da rede ainda chegará a um consenso no próximo bloco;
- **Consultas.** Escreve e executa qualquer consulta MongoDB para pesquisar o conteúdo de todas as transações, ativos, metadados e blocos armazenados. Funcionalidade desenvolvida pelo próprio MongoDB;
- **Suporte nativo de múltiplos ativos.** Sem moeda nativa no BigchainDB, qualquer ativo, token ou moeda pode ser emitido;
- **Baixa latência.** Uma rede global leva cerca de um segundo para chegar a um consenso sobre um novo bloco. Em outras palavras, a finalização da transação acontece rapidamente;
- **Customizável.** Projeta sua própria rede privada com a personalização de ativos, transações, permissões e transparência;
- **Permissão rica.** Permite a definição de permissões no nível da transação para garantir uma separação clara de funções e impor acesso seletivo;
- **Código aberto.** Código aberto para a comunidade para que todos possam usá-lo e construir seus próprios aplicativos a partir dele;
- **Público ou Privado.** Existe a possibilidade de implementar suas próprias redes públicas ou privadas para casos de uso específicos da indústria.

Apesar do BigchainDB mostrar que pode fornecer um desempenho mais alto do que as soluções de Blockchain nativas, está constantemente sendo criticado por não fornecer as mesmas garantias de confiança e modelo de tolerância a falhas que as Blockchains nativos [El-Hindi et al. 2019].

3. Metodologia

Um objetivo de pesquisa, normalmente, comporta uma ou mais hipóteses de trabalho. Segundo Wazlawick (2009) um bom objetivo de pesquisa, normalmente, terá a forma de demonstrar que a hipótese elaborada é verdadeira. Para alcançar tal objetivo nesta pesquisa, em seus aspectos metodológicos, envolveram-se os passos discutidos por Wazlawick (2009): i) elaborar um tema de pesquisa que determina uma área de conhecimento na qual se deseja trabalhar; ii) realizar uma revisão bibliográfica ou revisão do estado da arte; e iii) definir o objetivo da pesquisa. Esta primeira etapa da metodologia foi de grande importância na identificação do nível de relevância do tema deste ensaio. Além disso, ajudou no estabelecimento do escopo da pesquisa por meio das hipóteses e dos objetivos geral e específicos.

Em termos de caracterização, a presente pesquisa se comporta como pesquisa exploratória e não-experimental. Pesquisas exploratórias têm como objetivo principal o aprimoramento de ideias ou a descoberta de intuições [Gil 2002]. Segundo Gil (2002) grande parte das pesquisas exploratórias envolvem: i) levantamento bibliográfico; ii) entrevistas com pessoas que tiveram experiências práticas com o problema pesquisado; e iii) análise de exemplos que estimulem a compreensão. Neste ensaio, a natureza da pesquisa exploratória será utilizada para apresentar os conceitos básicos necessários para que se possa entender o problema e o objetivo da pesquisa, além de descrever, com maiores detalhes, as soluções de persistência de dados escolhidas e os motivos das suas escolhas. Portanto, a segunda etapa da metodologia se serviu da pesquisa exploratória para reunir

todo arcabouço teórico e os conhecimentos necessários para o entendimento do presente estudo.

Já que também é uma pesquisa não-experimental são aplicadas observações no intuito de tirar conclusões a partir de um arcabouço teórico reunido [Wazlawick 2009]. Em seguida, será detalhado o método de pesquisa, onde é especificada uma sequência de passos necessários para demonstrar que o objetivo proposto foi atingido [Wazlawick 2009]. Nesta artigo, a pesquisa de natureza não-experimental consiste no estudo sem a intervenção sistemática do pesquisador, observando e tirando conclusões a partir de um arcabouço teórico reunido [Wazlawick 2009]. Portanto, a pesquisa não-experimental, neste trabalho, compreende as etapas da metodologia que refletem a comparação e discussão das soluções de persistência de dados elencadas por meio das informações obtidas no levantamento bibliográfico na parte exploratória da pesquisa. Assim, a terceira e última etapa da metodologia, presente neste ensaio, utilizou a pesquisa não-experimental para apresentar os resultados obtidos e a discussão realizada.

4. Resultados e Discussão

Nesta seção foi feita uma discussão teórica, apontando alguns aspectos técnicos, por meio de comentários das aplicações orientadas a dados, de diferentes domínios, que foram elencadas para conduzir a discussão. Foram utilizadas aplicações de domínios diferentes pelo fato delas serem implementadas por requisitos de dados diferentes. Para cada aplicação elencada, serão mostrados seus requisitos de dados e, assim, confrontar com as características dos Bancos de Dados NoSQL e Bancos de Dados que implementam características de Blockchain. Sempre que possível, ampliar as discussões técnicas por meio das implementações MongoDB e BigchainDB.

4.1. Saúde: Aplicações de Análises Clínicas

Geralmente, aplicações desenvolvidas para análises clínicas, não necessitam de uma computação de alto desempenho e produzem muitos dados a serem processados. Dado os requisitos de dados de uma aplicação de análises clínicas, Blockchain pode ser uma ótima solução para armazenamento de dados, garantindo aspectos de segurança, confiabilidade e auditabilidade de algumas categorias de dados. Ao adotar Blockchain como um Banco de Dados para uma aplicação de análises clínicas, pode-se ter os seguintes benefícios: i) um médico não pode negar que emitiu um determinado laudo médico para um determinado paciente, pois Blockchain garante os aspectos de irrefutabilidade; ii) os dados presentes em um determinado laudo médico não podem sofrer alterações, pois Blockchain garante os aspectos de imutabilidade. Se unirmos os aspectos de imutabilidade e irrefutabilidade, temos uma maior confiabilidade nos dados presentes em uma rede Blockchain e ainda favorece os aspectos de auditabilidade; iii) outros aspectos que podem ser benéficos são os de privacidade e anonimidade, pois parte dos dados podem ser anonimizados em uma transação, favorecendo a proteção de dados sensíveis diante dos acessos.

No entanto, se observamos uma aplicação de análises clínicas em outra perspectiva, alguns requisitos de dados poderiam ser prejudicados diante da adoção da Blockchain como Banco de Dados. Suponha que desejamos alterar dados pessoais de um médico, pois um médico mudou de endereço ou telefone. Observe que os aspectos de imutabilidade poderiam gerar um maior problema de implementação. Se os dados do médico fossem

mantidos em uma Blockchain, teria-se que implementar, no nível de aplicação, um versionamento dos dados e apontar qual é o mais recente. Assim, ocorreria um aumento no tempo de desenvolvimento das aplicações, pois o desenvolvedor estaria preocupado com as regras de negócio e também com a complexidade do armazenamento diante de alguns requisitos de dados. Portanto, é muito importante avaliar o impacto da realização dos requisitos de dados nas soluções de gestão de dados. Como visto, algumas soluções podem beneficiar as aplicações orientadas a dados, mas podem também agregar esforços adicionais na etapa de implementação. Neste cenário de aplicações de análises clínicas orientadas a dados, uma solução híbrida poderia ser mais adequada, onde parte dos dados ficaria em uma Blockchain e outra parte seria gerenciada por um Banco de Dados NoSQL.

Diante disso, uma aplicação de análise clínica orientada a dados poderia ser implementada com o MongoDB e com o BigchainDB, utilizando o particionamento dos seus dados. Os dados de cadastros em geral, por exemplos cadastros dos médicos, outros profissionais e clínica, poderiam ser gerenciados no MongoDB, aproveitando sua características de facilidade com gestão de dados em JSON. Já os dados produzidos durante a execução de exames clínicos, poderiam ser armazenados em um BigchainDB. Dados de laudos médicos, profissionais que fizeram as coletas para exames e consultas realizadas poderiam se servir, por exemplo das características de imutabilidade do BigchainDB. Além disso, BigchainDB se serve do MongoDB, assim consultas em uma solução de armazenamento podem também ser empregadas na solução complementar, facilitando a troca de dados ou informações entre elas.

4.2. Aplicações de Internet das Coisas

Uma aplicação de Internet das Coisas, de um modo geral, realiza a conexão entre objetos físicos por meio da Internet. Os objetos físicos podem atuar como captadores de dados e fazem trocas de informações por meio da rede, em prol da aplicação que foi implementada servindo-se da Internet das Coisas. São exemplos de objetos físicos no âmbito da Internet das Coisas: *smartwatches*, computadores, *tablets*, *smart TVs*, lâmpada inteligente, caixa de som inteligente, tecnologias vestíveis etc. Se pensarmos no desenvolvimento de aplicações onde se tem diversos objetos físicos produzindo grandes volumes de dados a serem processados, então o Banco de Dados NoSQL é uma solução favorável para esse contexto, alinhando-se ao conceito de Big Data. Na Seção 2 foi comentado que Bancos de Dados NoSQL possuem características que possibilitam processamento em larga escala, pois possuem estruturas flexíveis, facilidade de expansão e realizam leituras e escritas de dados eficientemente. As estruturas flexíveis ajudam no processo de estruturação dos dados a serem processados, pois os objetos físicos não possuem estruturas de dados uniformes. A facilidade de expansão impacta no aumento, sem muito esforço, do poder computacional da infraestrutura de gestão de dados, podendo acelerar o processamento de grandes volumes de dados em larga escala e também melhorar a disponibilidade da aplicação. A realização de leituras e escritas com alta velocidade também impacta no desempenho total do processamento dos grandes volumes de dados.

Entretanto, se observamos pelos aspectos de segurança e privacidade, em alguns contextos de aplicações de Internet das Coisas, um Banco de Dados NoSQL pode não ser uma alternativa adequada. Por exemplo, considere que numa aplicação de Internet das Coisas, um dos requisitos de dados atende à possibilidade de auditar, de forma confiável,

os dados que foram processados e trafegados entre os objetos físicos. Outro requisito de dados pode consistir na possibilidade de controlar a identidade dos objetos físicos, que coletam dados, para fins de rastreabilidade e auditabilidade. Além disso, controlar a entrada e saída de objetos físicos por meio dos mecanismos de permissão da Blockchain, evita a entrada de objetos maliciosos na rede. Algumas aplicações deste domínio podem também requisitar aspectos de imutabilidade dos dados para fins de histórico do que foi processado. A característica de irrefutabilidade pode ser interessante no cenário onde há a necessidade de identificar o objeto físico autor de tal processamento. Portanto, dependendo dos cenários e dos requisitos de dados em um mesmo domínio de aplicação orientada a dados, uma solução de gestão de dados pode ter vantagens e desvantagens.

Diante disso, uma aplicação de Internet das Coisas orientada a dados, dependendo dos seus requisitos de dados, podem se servir tanto de Banco de Dados NoSQL quanto Blockchain. Assim, aplicações que necessitam de processamento de grandes volumes de dados em alta escala, onde o desempenho é essencial, os dados devem ser manipulados por um MongoDB. O armazenamento de dados em um BigchainDB seria viável em aplicações onde seriam importantes as características de imutabilidade, irrefutabilidade, segurança e auditoria.

5. Conclusão

Este artigo apresentou um estudo exploratório inicial para suportar futuros aprofundamentos de análises comparativas entre Banco de Dados NoSQL e armazenamento de dados em Blockchain, destacando benefícios às aplicações orientadas a dados. Para isso, utilizou-se os aspectos teóricos sobre Banco de Dados NoSQL e armazenamento de dados em Blockchain. Além disso, também comentou-se sobre as soluções MongoDB, como Banco de Dados NoSQL, e o BigchainDB, como Banco de Dados que implementa características de Blockchain. Em termos metodológicos, este ensaio serviu-se da pesquisa exploratória e da pesquisa não-experimental. Na etapa da pesquisa exploratória, em particular, foi utilizada o levantamento bibliográfico para reunir os conceitos teóricos necessários para o entendimento do trabalho. Já a pesquisa não-experimental foi adotada para tirar conclusões a partir de um arcabouço teórico reunido na etapa da pesquisa exploratória.

Como resultados e discussão desta pesquisa foram identificados como os requisitos de dados, de domínios de aplicações orientadas a dados diferentes, podem se beneficiar de Banco de Dados NoSQL e/ou armazenamento de dados em Blockchain. Para confrontar os aspectos teóricos com os aspectos técnicos, foram utilizados, como soluções de implementação, o Banco de Dados NoSQL MongoDB e o BigchainDB que é um Banco de Dados que possui características de Blockchain. Diante dos resultados e discussões, pode-se perceber que foi possível responder à questão do problema de pesquisa definida na Seção 1. Banco de Dados NoSQL e o armazenamento de dados em Blockchain podem beneficiar de classes distintas de aplicações orientadas a dados, por exemplo, uma aplicação que emite laudos médicos pode se servir das características de imutabilidade e irrefutabilidade do armazenamento em Blockchain para garantir a autoria real do laudo médico e que esse laudo médico não sofra alterações, aumentando a segurança e confiabilidade dos dados. Já o uso de algoritmos de consenso das Blockchains para aplicações que necessitam de requisitos de dados com processamento de alto desempenho, como aqueles comentados na Seção 2, reduzem o desempenho das aplicações na ação de validar um novo bloco na rede. Além disso, Blockchains são ineficientes para aplicações que neces-

sitam implementar ações de *Create, Read, Update e Delete (CRUD)*, pois as características de imutabilidade dos dados não colaboram com tal necessidade.

Ao final deste estudo, foi possível confirmar as duas hipóteses que foram formuladas: i) durante o estudo exploratório, levantou-se que ambas soluções de persistência possuem características diferentes quanto à gestão de dados, por exemplo Blockchain tem dados imutáveis e desempenho inferior a outra solução; e ii) a própria análise comparativa realizada na Seção 4 e os comentários no parágrafo anterior, revelam aspectos de vantagens e desvantagens em relação aos requisitos e da natureza das aplicações orientadas a dados.

Nos aspectos de limitações e ameaças à validade do estudo, pode-se destacar:

- O estudo foi conduzido seguindo a pesquisa exploratória e a pesquisa não-experimental, logo, todo o conhecimento produzido foi obtido por meio da observação, sem intervenção sistemática do pesquisador, e realizando conclusões a partir de um arcabouço teórico reunido. Assim, não existe uma formalidade científica que mensure, na prática, os resultados e a discussão. Portanto, práticas por meio de experimentos, provas de conceito ou protótipos, poderiam reforçar as hipóteses e a questão do problema científico;
- Outras soluções de persistência, de fabricantes diferentes, poderiam impactar nos resultados e na discussão. Neste trabalho só foi discutido o MongoDB e o BlockchainDB como soluções de persistência. Assim, seria interessante uma exploração maior de soluções de persistência de outros fabricantes e realizar análises de desempenho diante das aplicações e seus requisitos de dados;
- Somente duas aplicações foram discutidas na Seção 4, o ideal é reunir mais aplicações, elencar seus requisitos de dados e fazer um comparativo com os recursos das soluções de persistência.

Como trabalhos futuros, almeja-se: i) realizar um estudo mais aprofundado com relação à análise comparativa inicial que foi feita neste estudo; ii) fazer análises comparativas, também na perspectiva dos benefícios das aplicações orientadas a dados, com outras soluções de gestão de dados, como por exemplo os Bancos de Dados NewSQL; e iii) construir protótipos ou provas de conceito para validar, com mais aspectos técnicos, as análises comparativas realizadas neste ensaio, pois os resultados obtidos e discussão foram na perspectiva teórica.

6. Agradecimentos

Este estudo foi parcialmente financiado pela Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP n.º 08910797/2020).

Referências

- Aleksieva, V., Valchanov, H., and Huliyan, A. (2020). Smart contracts based on private and public blockchains for the purpose of insurance services. In *2020 International Conference Automatics and Informatics (ICAI)*, pages 1–4.
- Andreoli, R., Cucinotta, T., and Pedreschi, D. (2021). RT-MongoDB: A NoSQL Database with Differentiated Performance. In Helfert, M., Ferguson, D., and Pahl, C., editors, *Proceedings of the 11th International Conference on Cloud Computing and Services Science (CLOSER)*, pages 77–86. SciTePress.

- Behl, D., Kodeswaran, P., Ramakrishna, V., Sen, S., and Vinayagamurthy, D. (2020). Trusted data notifications from private blockchains. In *2020 IEEE International Conference on Blockchain (Blockchain)*, pages 53–61.
- BigchainDB (2021a). BigchainDB - The Blockchain Database. Disponível em: <https://www.bigchaindb.com/>. Acessado em 06 de junho de 2021.
- BigchainDB (2021b). Features & Use Cases - BigchainDB. Disponível em: <https://www.bigchaindb.com/features/>. Acessado em 12 de junho de 2021.
- Braga, A., Marino, F., and Santos, R. (2017). *Segurança de Aplicações Blockchain Além das Criptomoedas*, chapter 3, pages 99–148. SBC. Minicursos do XVII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSEG).
- Chen, W., Xu, Z., Shi, S., Zhao, Y., and Zhao, J. (2018). A survey of blockchain applications in different domains. In *Proceedings of the 2018 International Conference on Blockchain Technology and Application, ICBTA 2018*, page 17–21, New York, NY, USA. Association for Computing Machinery.
- El-Hindi, M., Binnig, C., Arasu, A., Kossmann, D., and Ramamurthy, R. (2019). BlockchainDB: A Shared Database on Blockchains. *Proc. VLDB Endow.*, 12(11):1597–1609.
- Elmasri, R. and Navathe, S. B. (2011). *Sistemas de Banco de Dados*. Pearson Education Brasil, 6a. edition.
- Fan, C., Ghaemi, S., Khazaei, H., and Musilek, P. (2020). Performance evaluation of blockchain systems: A systematic survey. *IEEE Access*, 8:126927–126950.
- Farias, F. A. d. M. (2014). Avaliação de Desempenho entre Bancos de Dados Relacionais e NoSQL. Monografia de Graduação. Bacharel em Sistemas de Informação, Centro de Ciências Aplicadas à Educação, Universidade Federal da Paraíba, Rio Tinto, Brasil.
- Freitas, M. C., Souza, D. Y., and Salgado, A. C. (2015). Mapeamentos conceituais entre os modelos relacional e nosql: Uma abordagem comparativa. *Revista Principia - Divulgação Científica e Tecnológica do IFPB*, 1(28):37–50.
- Gil, A. C. (2002). *Como Elaborar Projetos de Pesquisa*. Atlas, São Paulo, 4a. edition.
- Greve, F., Sampaio, L., Abijaude, J., Coutinho, A., Valcy, I., and Queiroz, S. (2018). *Blockchain e a Revolução do Consenso sob Demanda*, chapter 5, pages 1–52. SBC. Minicursos do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC).
- Han, J., E, H., Le, G., and Du, J. (2011). Survey on nosql database. In *2011 6th International Conference on Pervasive Computing and Applications*, pages 363–366.
- Marinho, C. S. S., Costa Filho, J. S., Moreira, L. O., and Machado, J. C. (2020). Using a Hybrid Approach to Data Management in Relational Database and Blockchain: a Case Study on The E-health Domain. In *2020 IEEE International Conference on Software Architecture Companion (ICSA-C)*, pages 114–121.
- MongoDB (2021a). JSON and BSON - MongoDB. Disponível em: <https://www.mongodb.com/json-and-bson/>. Acessado em 11 de junho de 2021.

- MongoDB (2021b). The most popular database for modern apps - MongoDB. Disponível em: <https://www.mongodb.com/>. Acessado em 06 de junho de 2021.
- Moreira, L. O. (2014). *Abordagem para Qualidade de Serviço em Banco de Dados Multi-Inquilinos em Nuvem*. PhD thesis, Doutorado em Ciência da Computação, Centro de Ciências, Universidade Federal do Ceará, Fortaleza, Brasil.
- Moreira Neto, M., Marinho, C. S. S., Coutinho, E. F., Moreira, L. O., Machado, J. C., and Souza, J. N. (2020). Research Opportunities for E-health Applications with DNA Sequence Data using Blockchain Technology. In *2020 IEEE International Conference on Software Architecture Companion (ICSA-C)*, pages 95–102.
- Niya, S. R., Schiller, E., Cepilov, I., Maddaloni, F., Aydinli, K., Surbeck, T., Bocek, T., and Stiller, B. (2019). Adaptation of proof-of-stake-based blockchains for iot data streams. In *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pages 15–16.
- Nofer, M., Gomber, P., Hinz, O., and Schiereck, D. (2017). Blockchain. *Business and Information Systems Engineering*, 59(3):183–187.
- Rubio, F., ., P. V., and Reyes Ch, R. P. (2020). Nosql vs. sql in big data management: An empirical study. *KnE Engineering*, 5(1):40–49.
- Sadalage, P. J. and Fowler, M. (2013). *NoSQL Essencial: Um Guia Conciso para o Mundo Emergente da Persistência Poliglota*. Novatec.
- Soares, B. E. and Boscarioli, C. (2013). Modelo de banco de dados colunar: Características, aplicações e exemplos de sistemas. In *IX Escola Regional de Banco de Dados (ERBD)*, pages 1–10.
- Wazlawick, R. S. (2009). *Metodologia de Pesquisa para Ciência da Computação*. Elsevier Editora, Rio de Janeiro, 1a. edition.
- Yaga, D. and Mell, P. (2018). *NISTIR 8202: Blockchain Technology Overview*, chapter 1, pages 1–68. National Institute of Standards and Technology.
- Yang, X., Chen, Y., and Chen, X. (2019). Effective scheme against 51% attack on proof-of-work blockchain with history weighted information. In *2019 IEEE International Conference on Blockchain (Blockchain)*, pages 261–265.