



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
SISTEMAS DE INFORMAÇÃO

Pedro Henrique Dias Nobrega

Big data e Blockchain: um sistema para análise de dados educacionais.

Florianópolis
2023

Pedro Henrique Dias Nobrega

Big data e Blockchain: um sistema para análise de dados educacionais.

Trabalho de conclusão de curso submetido ao Curso de Graduação em Sistemas de Informação da Universidade Federal de Santa Catarina como requisito para a obtenção do título Bacharel em Sistemas de Informação.

Orientadora: Giovana Nunes Inocência, Tecga.

Coorientador: Prof. Jean Everson Martina, Dr.

Florianópolis

2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Nobrega, Pedro Henrique Dias

Big data e Blockchain : um sistema para análise de dados educacionais. / Pedro Henrique Dias Nobrega ; orientadora, Giovana Nunes Inocêncio, coorientador, Jean Everson Martina, 2023.

71 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Sistemas de Informação, Florianópolis, 2023.

Inclui referências.

1. Sistemas de Informação. 2. Big Data. 3. Blockchain. 4. Gestão educacional. I. Inocêncio, Giovana Nunes. II. Martina, Jean Everson. III. Universidade Federal de Santa Catarina. Graduação em Sistemas de Informação. IV. Título.

Pedro Henrique Dias Nobrega

Big data e Blockchain: um sistema para análise de dados educacionais.

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de Bacharel em Sistemas de Informação e aprovado em sua forma final pelo Curso de Graduação de Graduação em Sistemas de Informação

Florianópolis, 2023

Prof. Álvaro Junio Pereira Franco, Dr.
Coordenador do Curso

Banca Examinadora:

Giovana Nunes Inocência, Tecga.
Orientadora
Universidade Federal de Santa Catarina

Prof. Jean Everson Martina, Dr.
Co-orientador
Universidade Federal de Santa Catarina

Lucas Machado da Palma, Me.
Universidade Federal de Santa Catarina

Wellington Fernandes Silvano, Me.
Universidade Federal de Santa Catarina

Este trabalho é dedicado à todos
que me auxiliaram a superar
todos os desafios.

RESUMO

Com a descentralização do sistema educacional de ensino, distribuído em todo o país, observa-se a necessidade de uma consulta unificada dos dados a nível nacional. Por causa da carência de união dos dados e da própria natureza das informações, surge a proposta de unificar estes dados em uma Blockchain, trazendo padronização e controle rígido aos dados dos participantes. Pode-se definir o termo Blockchain como um livro-razão implementado de uma maneira distribuída, geralmente sem uma autoridade central. Tendo sido originalmente cunhado por Satoshi Nakamoto — pseudônimo do suposto criador do Bitcoin— a Blockchain tem como função registrar dados de forma resistente ao tempo e a possíveis tentativas de modificação dos mesmos, usando mecanismo de consenso para registrar novos blocos no livro-razão distribuído. Já com os dados unificados e disponíveis, é possível utilizar um sistema de consulta aos mesmos por meio de alguma ferramenta, que suporte uma grande quantidade de dados, um posterior incremento dessas informações e que consiga conversar com a Blockchain utilizada. Essa dificuldade de análise faz com que seja primordial o uso de Big Data, que é o termo que define conjuntos de dados massivos que possuem uma grande variedade e estruturas complexas com dificuldades de ordenamento, análise e visualização para futuros processamentos ou resultados. Com uma grande quantidade de informações disponíveis, considera-se a necessidade de fazer análises, modelos e explorações nesses conjuntos de dados. Com isso, surge a área de Big Data, focada em fornecer uma maneira eficiente e eficaz de manipular e armazenar esses dados. Utilizando ferramentas como o Metabase, desenvolvida especificamente para análise de grandes dados, é possível ter uma visão analítica das informações educacionais. Dessa forma, é possível garantir uma maior gestão educacional sobre o ensino brasileiro, que possibilitaria um sistema de decisão mais preciso e conciso. Com isso foi criado um sistema para visualização desses dados em formato de dashboards, apresentando painéis variados, que poderão ser utilizados para aperfeiçoar o processo de tomada de decisões dentro do Ministério da Educação e pelo Ministro da Educação.

Palavras-chave: Blockchain; Big Data; Business Intelligence; Gestão Educacional

ABSTRACT

With the decentralization of the educational system, distributed throughout the country, there is a need for a unified consultation of data at the national level. Due to the lack of data unity and the very nature of the information, there is a proposal to unify these data on a Blockchain, bringing standardization and strict control to the participants' data. The term Blockchain can be defined as a ledger implemented in a distributed way, usually without a central authority. Originally coined by Satoshi Nakamoto — the pseudonym of the supposed creator of Bitcoin — the Blockchain's function is to record data in a way that is resistant to time and potential attempts to modify it, using a consensus mechanism to record new blocks in the distributed ledger. With the data unified and available, it is possible to use a system to consult them through some tool, which supports a large amount of data, a subsequent increase of this information, and that can communicate with the used Blockchain. This difficulty in analysis makes it paramount to use Big Data, which is the term that defines massive data sets that have a great variety and complex structures with difficulties in ordering, analyzing, and visualizing for future processing or results. With a large amount of information available, there is a need to perform analyses, models, and explorations in these data sets. Thus, the field of Big Data emerges, focused on providing an efficient and effective way to manipulate and store these data. Using tools like Metabase, specifically developed for the analysis of large data, it is possible to have an analytical view of educational information. In this way, it is possible to ensure greater educational management over Brazilian education, which would allow for a more precise and concise decision-making system. With this, a system was created for the visualization of these data in the form of dashboards, presenting various panels, which can be used to improve the decision-making process within the Ministry of Education and by the Minister of Education.

Keywords: Blockchain; Big Data; Business Intelligence; Educational Management

LISTA DE FIGURAS

Figura 1 - Criptografia de Chave Pública.....	21
Figura 2 - Representação de uma rede computacional distribuída.....	23
Figura 3 - Representação da blockchain.....	25
Figura 4 - Fluxograma de ingestão dos dados.....	38
Figura 5 - Alunos por sexo e idade.....	42
Figura 6 - Nota média por disciplina.....	43
Figura 7 - Desempenho por estado de origem.....	44
Figura 8 - Nota média das disciplinas por período.....	45

SUMÁRIO

1. INTRODUÇÃO.....	15
1.1. JUSTIFICATIVA.....	16
1.2. OBJETIVOS.....	17
1.2.1. Objetivo geral.....	17
1.2.2. Objetivos específicos.....	17
1.3. METODOLOGIA DE PESQUISA.....	18
2. FUNDAMENTAÇÃO TEÓRICA.....	20
2.1. CRIPTOGRAFIA ASSIMÉTRICA.....	20
2.2. COMPUTAÇÃO DISTRIBUÍDA.....	22
2.3. BLOCKCHAIN.....	24
2.4. SMART CONTRACTS.....	26
2.5. HYPERLEDGER FABRIC.....	27
2.6. BIG DATA.....	28
2.7. METABASE.....	29
3. METODOLOGIA.....	32
4. DESENVOLVIMENTO.....	34
4.2.1. Público alvo.....	37
4.2.2. Protótipo.....	37
5. RESULTADOS E DISCUSSÃO.....	42
5.1. CONSIDERAÇÕES FINAIS.....	46
6. CONCLUSÃO.....	48
REFERÊNCIAS.....	50
APÊNDICE A - ARTIGO DA MONOGRAFIA.....	53

1. INTRODUÇÃO

O Brasil possui um sistema educacional descentralizado, apresentando 2574 universidades espalhadas pelo país, fazendo com que existam diversos registros dispersos do desempenho acadêmico dos alunos, assim como dados demográficos e de permanência, nos cursos de graduação, que dificultam uma possível observação analítica sobre o sistema como um todo (INEP, 2021). Com o objetivo de tornar a abertura de processos regulatórios da educação superior no Brasil, em 2007 foi criado o e-MEC, visando dar agilidade e eficiência à comunicação entre o Ministério da Educação (MEC) e as instituições de ensino superior (IES).

Com base na dificuldade enfrentada pelo MEC, no artigo “*Blockchain and smart contracts for higher education registry in Brazil*”, os autores propuseram a criação de um novo modelo que faz uso de tecnologias de *blockchain*, e da natureza do próprio sistema para criar um mecanismo de consenso e padronização dos dados. Podendo ser definido como um livro-razão digital implementado de uma maneira distribuída, geralmente sem uma autoridade central, a tecnologia *blockchain* com ferramentas de *big data* pode ser uma solução para o problema da informação dispersa (YAGA, MELL, ROBY E SCARFONE, 2019).

Como solução ao problema dos dados despadronizados e sem coexistência, nasceu o Jornada do Estudante, que através de uma *blockchain* privada, uma *blockchain* onde os participantes podem ser escolhidos e as transações são visíveis apenas por membros credenciados, criou padrões para o armazenamento de dados e trouxe uma fonte de dados comum, sendo estes as próprias transações dentro da *blockchain*. Assim, o projeto Jornada do Estudante trouxe um novo meio de acompanhamento dos dados estudantis do Brasil, unificando-os em uma única plataforma que pode ser consultada por membros credenciados.

O termo *big data* define conjuntos de dados massivos que possuem uma grande variedade e estruturas complexas com dificuldades de ordenamento, análise e visualização para futuros processamentos ou resultados destinados a facilitar que um usuário com conhecimentos de análise de dados consiga navegar pelo universo de dados diversos, incluindo dados provenientes de *blockchains*.

Buscando mecanismos de decisão mais precisos e concisos, surge a necessidade de criar uma integração entre soluções de *big data* e *blockchain* possibilitando uma visão analítica dos dados. Ferramentas de *big data*, como por

exemplo o *Metabase*, são uma maneira *open-source*, ou seja, de código aberto, de facilitar a análise de informações. A *blockchain* privada proposta faz uso do *Hyperledger Fabric*, que é uma plataforma para soluções baseadas em livro-razão distribuído, sustentada por uma arquitetura modular que entrega um alto grau de confidencialidade, resiliência, flexibilidade e escalabilidade e foi projetada para suportar implementações para diversos componentes.

Pela natureza educacional da *blockchain* analisada durante esta pesquisa, foi constatado que a gestão educacional é uma etapa essencial na implementação do projeto de integração de *big data* e *blockchain* do Jornada do Estudante, pois está diretamente relacionada com a obtenção de dados educacionais de instituições de ensino, e pode ser definida como o reconhecimento da importância da participação consciente e esclarecida das pessoas nas decisões sobre a orientação e planejamento de seu trabalho (LÜCK, Heloísa).

Diante do problema dos dados descentralizados, da falta de padrão dentre as universidades, e buscando fornecer uma melhor gestão educacional, esta pesquisa busca solucionar o processo de conexão entre esses dois universos, propondo uma maneira de consultar os dados sequenciais padronizados da *blockchain* da Jornada do Estudante de forma eficiente e eficaz, trazendo *big data* e análise de dados ao mundo educacional.

1.1. JUSTIFICATIVA

Este projeto busca analisar dados educacionais, auxiliar na tomada de decisão, a partir de informações do desempenho dos alunos com base em diversos fatores, como o estado de origem do estudante e sua idade. Com isso, o processo de decisões internas de educação do Brasil será aperfeiçoado, com base em dados reais e coesos sobre o desenvolvimento acadêmico dos alunos da rede pública. Com essas informações em mãos, será possível moldar e estabelecer padrões acelerando a evolução do sistema educacional.

De acordo com o *Statista*, 2023, a *blockchain* tem apresentado um crescimento exponencial, com um valor de mercado previsto de 40 bilhões de dólares em 2025, fazendo com que o tema tenha um desenvolvimento constante, surgindo novas soluções a todo momento. Com isso surge a necessidade de analisar os dados gerados para extrair valor desse meio e permitir aos servidores do

Ministério da Educação (MEC) uma visão sistemática do progresso estudantil, fornecendo mais uma fonte de informação no processo de tomada de decisão.

Conforme os dados da Demografia Médica no Brasil, 2020, “O Brasil deixou de formar, em cinco anos, quase 36 mil médicos especialistas, considerando as vagas autorizadas, mas não ocupadas.”. Essa taxa de vacância poderia ser mitigada fazendo uso da solução proposta neste projeto, onde seria possível analisar, por exemplo, a real quantidade de residentes que entrariam no processo seletivo e realocar as vagas conforme a necessidade. Essa lacuna de informação faz com que profissionais que desejam se especializar e desempenhar um papel importante na sociedade, não consigam vagas em suas áreas. Com a solução proposta e um cruzamento de dados com informações do Ministério da Saúde, seria possível prever essas oportunidades e preparar a oferta de vagas para suportar essa demanda.

1.2. OBJETIVOS

Para o desenvolvimento desta pesquisa foram definidos os seguintes objetivos: geral e específicos.

1.2.1. Objetivo geral

Esta pesquisa tem como objetivo geral a análise dos dados educacionais da rede *blockchain* da Jornada do Estudante através de uma interface de consulta de *big data*.

1.2.2. Objetivos específicos

- Conectar uma ferramenta analítica, que fornecerá a plataforma de análise de dados, à *blockchain*;
- Modelar os dados de uma *blockchain* a fim de satisfazer as necessidades específicas relacionadas à gestão educacional;
- Construir um *dashboard* com exemplos visuais de informações possíveis a serem obtidas com os dados da *blockchain*;

1.3. METODOLOGIA DE PESQUISA

A pesquisa possui abordagem qualitativa, em que a primeira etapa consistiu em leituras exploratórias do tema, buscando o que está sendo desenvolvido na área e identificando o estado da arte, extraindo os conceitos de *blockchain*, assim como obter o estado da arte de *big data* e ferramentas analíticas em processos de gestão de conhecimento. A partir disso, é feita uma observação a fim de identificar como as tecnologias propostas na pesquisa estão sendo utilizadas pela comunidade, com aprofundamento em como adaptá-las para as necessidades da pesquisa.

Com os resultados qualitativos, foi realizada uma análise quantitativa com objetivo de extrair métricas relacionadas às ferramentas selecionadas, de forma a fundamentar a escolha da mesma ou expandir o leque de opções para novas ferramentas, descartando as que possuem performance abaixo do esperado ou as que possuem custos acima do esperado para a solução proposta.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, o leitor será introduzido aos conceitos base para entendimento do funcionamento das tecnologias utilizadas no desenvolvimento do trabalho, assim como os motivos dessas escolhas.

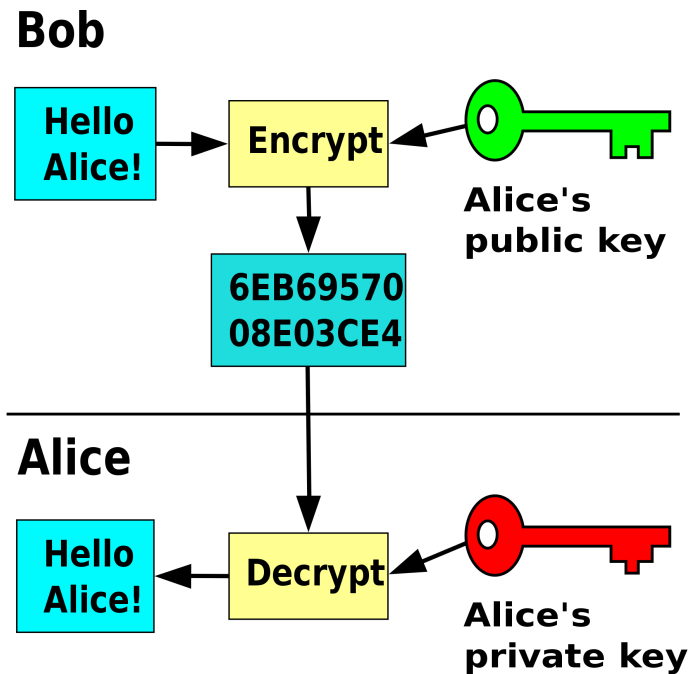
2.1. CRIPTOGRAFIA ASSIMÉTRICA

O termo criptografia define um conjunto de técnicas que visam transformar uma mensagem de modo a torná-la incompreensível para quem não esteja autorizado a lê-la. Essa transformação é realizada por meio de um algoritmo, que é um conjunto de regras matemáticas que alteram a forma da mensagem (NETO, 2002)

A Criptografia Assimétrica, também conhecida como criptografia de chave pública, é uma técnica de criptografia que utiliza um par de chaves (uma pública e uma privada) que são matematicamente relacionadas, porém não idênticas. Assim, tudo que a chave privada criptografa a pública descriptografa, e vice-versa.

A chave pública é usada para criptografar mensagens, enquanto a chave privada é usada para descriptografá-las (SCHNEIER, 1996), fornecendo sigilo às informações criptografadas, assim como o inverso, utilizando a chave privada para criptografar mensagens e a pública para descriptografá-las, fornecendo autenticidade às informações. Essa forma de criptografia proporciona uma segurança pela dificuldade de deduzir a chave privada a partir da chave pública. A Figura 1 representa como funciona esse tipo de criptografia na prática, onde Bob utiliza a chave pública de Alice para criptografar a mensagem, que apenas pode ser descriptografada pela chave privada de Alice.

Figura 1 - Criptografia de Chave Pública.



Fonte: (https://en.wikipedia.org/wiki/Public-key_cryptography)

Uma das formas de criptografia assimétrica amplamente utilizadas para proteger a confidencialidade, a integridade e a autenticidade de informações é o Rivest-Shamir-Adleman (RSA), que foi um dos primeiros sistemas propostos que fazem uso desse método, ainda mantendo a popularidade atualmente. O RSA baseia-se na dificuldade de fatorar grandes números primos, um problema para o qual não se conhece nenhuma solução eficiente (RIVEST, SHAMIR, ADLEMAN, 1978). Isso torna o RSA um sistema de criptografia assimétrica robusto e confiável.

A criptografia de chave pública possui um papel essencial na segurança da internet. Ela é utilizada em diversas aplicações de segurança, como em autenticação e integridade de dados. Por exemplo, os certificados digitais usados em conexões HTTPS, ou seja, uma conexão HTTP feita de forma segura, se baseia em criptografia assimétrica para garantir a identidade de um servidor *web* (RESCORLA, 2001). A criptografia assimétrica também é essencial para *blockchain*, onde cada integrante da rede possui um par de chaves pública e privada e podem, por

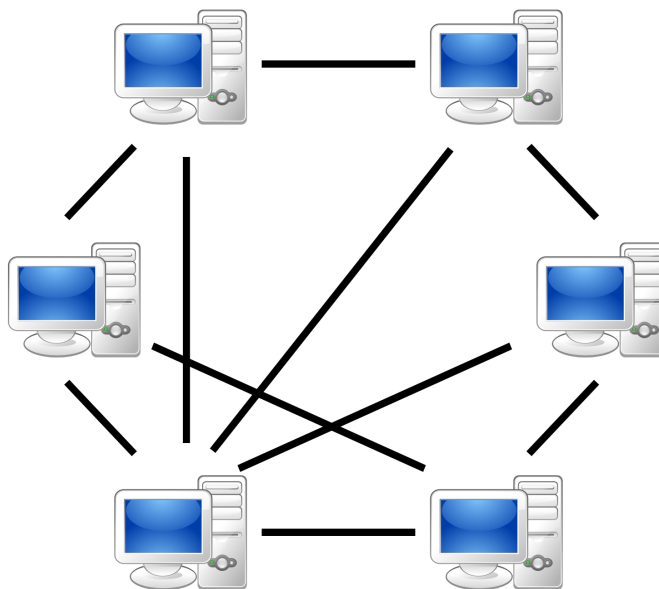
exemplo, utilizar suas chaves para assinar uma transação, a partir do uso da chave privada para criptografar a mesma, fornecendo autenticidade às transações da *blockchain*.

Porém, a criptografia assimétrica possui desvantagens em relação a outros métodos de criptografia. Ela é significativamente mais lenta do que a criptografia simétrica, e, por esse motivo, em muitos sistemas práticos a criptografia assimétrica é usada para trocar uma chave simétrica, que é então usada para criptografia de dados em massa (MENEZES, VAN OORSCHOT, VANSTONE, 1996). Apesar de seus desafios, a criptografia assimétrica continua sendo crucial para a segurança cibernética moderna.

2.2. COMPUTAÇÃO DISTRIBUÍDA

A Computação Distribuída se refere a um estilo de computação no qual o processamento é distribuído em vários nós em rede. Sistemas distribuídos são sistemas computacionais na qual componentes localizados em redes se comunicam e coordenam suas ações apenas através de trocas de mensagens (COULOURIS, DOLLIMORE, KINDBERG, 2005). Ao invés da centralização em um único computador, usando o paradigma padrão da computação, as informações e funções são distribuídas por diversos dispositivos.

Figura 2 - Representação de uma rede computacional distribuída.



Fonte: (<https://pt.wikipedia.org/wiki/Peer-to-peer>)

A Figura 2 representa o modelo distribuído, em que vários computadores independentes, chamados de nós, cada um com seu processamento, memória e armazenamento locais, comunicam-se com os outros nós através de um sistema de comunicação (TANENBAUM, VAN STEEN, 2007). Essa comunicação oferece um compartilhamento de recursos vantajoso, em que se pode compartilhar tanto dados quanto poder de processamento e armazenamento. Com essa vantagem, o processamento em várias máquinas pode proporcionar redundância, possibilitando que algumas máquinas da rede parem de funcionar sem que o sistema, como um todo, pare de funcionar, proporcionando um sistema mais resiliente a falhas. Essa redundância é essencial para uma *blockchain*, onde, por exemplo, os dados das transações estão distribuídos dentre os nós participantes e ficam disponíveis para todos os membros da rede.

Um aspecto-chave da computação distribuída é o *middleware*, uma camada de software que fornece um modelo de programação e oculta a heterogeneidade do *hardware* subjacente, assim não é necessário projetar o sistema para rodar em *hardware* específico, barateando o custo de implementação desses sistemas, bem como as diferenças nas linguagens de programação e no sistema operacional

(EMMERICH, 2000). Isso permite aos desenvolvedores uma abstração extra, tirando deles a responsabilidade de manusear o sistema distribuído e possibilitando a concentração apenas na inovação e criação de valor, excluindo a preocupação com os detalhes de baixo nível do sistema.

Apesar dessas vantagens, a computação distribuída não é isenta de desafios. Um dos principais é a coordenação entre nós, a fim de garantir o processamento correto das informações, e a garantia de consistência dos dados especialmente em situações em que várias operações acontecem simultaneamente, que podem gerar condições de corrida (LYNCH, 1996). Mesmo com esses novos problemas, com o surgimento da internet e da computação em nuvem, a computação distribuída se tornou um pilar central da computação moderna, possibilitando avanços significativos em áreas que demandam muito poder de processamento, como *big data*, aprendizado de máquina, internet das coisas e outros diversos aspectos da internet moderna.

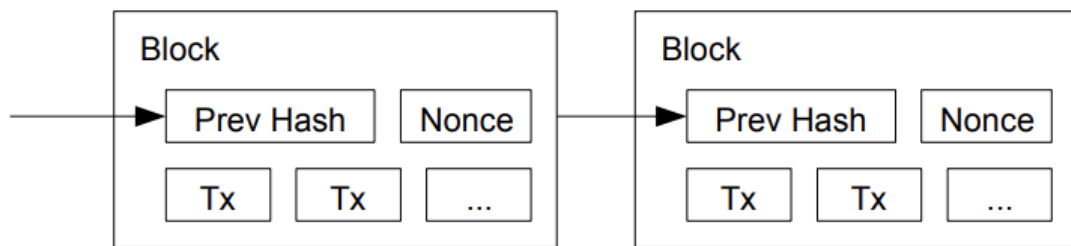
2.3. BLOCKCHAIN

A *Blockchain*, inicialmente popularizado como a tecnologia subjacente ao *Bitcoin*, é uma tecnologia de contabilidade distribuída que permite um registro seguro, transparente e resistente à manipulação de transações (NAKAMOTO, 2008). O termo se refere ao encadeamento de blocos de um livro-razão (*ledger*) e distribuído, na qual todas as transações de uma rede *blockchain* estão localizadas, podendo ter comportamento público, como no *Bitcoin*, onde todos podem ver todas as transações, ou privado, como no *Hyperledger Fabric*, onde se tem controle de quem pode contribuir para a rede e quem acessa cada informação da rede. A segurança e a imutabilidade da *blockchain* derivam do uso de criptografia de chave pública e do conceito de mineração ou prova de trabalho e da utilização de *hashes*, garantindo integridade, autenticidade e ordenamento à *blockchain*.

O termo *blockchain* traduz o próprio funcionamento da rede, em que cada bloco contém um *hash*, que é uma função matemática que mapeia dados de comprimento variável para dados de comprimento fixo, do bloco anterior, criando uma ligação contínua que torna quase impossível a alteração de blocos passados sem ser detectado (ANTONOPOULOS, 2014). Isso faz da *blockchain* uma plataforma robusta para transações seguras e transparentes. A Figura 3 é uma

representação visual de como a blockchain do *Bitcoin* encadeia os blocos, onde o bloco mais à direita possui uma referência ao bloco anterior, proporcionando a integridade da *blockchain*. É possível visualizar o "Nonce", que é um número arbitrário que só pode ser usado uma vez, servindo como fator de aleatoriedade para formação da hash do bloco da *blockchain* do *Bitcoin*.

Figura 3 - Representação da blockchain.



Fonte: (<https://bitcoin.org/bitcoin.pdf>)

Nesse contexto, cada bloco é uma informação de todas suas transações, armazenadas em formato de *hash*, representando a escrita (inserção à cadeia de blocos) das mesmas na rede, distribuindo essa nova confirmação para todos seus participantes. A rede é composta por uma série de computadores distribuídos, conhecidos e atuantes na rede, denominados de nós. Na rede do *Bitcoin*, esses nós são distribuídos entre nós validadores e nós mineradores.

A decisão de quais blocos devem ser inseridos na rede é realizada à partir de um mecanismo de consenso, em que cada nó validador realiza uma checagem na *hash* de cada transação para validar a não ocorrência de fraude, como por exemplo o "double spending", que, na rede *Bitcoin*, acontece quando um mesmo *Bitcoin* é transacionado para duas carteiras diferentes ao mesmo tempo, gerando novos *Bitcoins* no caminho e acarretando em uma fraude.

Os nós mineradores são responsáveis por adicionar mais blocos à rede, através de um processo denominado prova de trabalho, na qual, por meio de um sorteio computacional, cada nó recebe um quebra-cabeças matemático complexo e o primeiro que resolver detém o direito de adicionar o novo bloco na rede. Esses nós recebem uma recompensa ao adicionar novos blocos à rede (taxa de transação).

Com essa natureza descentralizada, a *blockchain* permite a verificação e registro de transações sem depender de uma autoridade central, possibilitando uma contribuição dos participantes para a obtenção de um consenso.

No entanto, o *Bitcoin* também enfrenta desafios e limitações. A escalabilidade é um dos problemas principais, pois a necessidade de todos os nós validarem todas as transações pode limitar a capacidade da rede. Além disso, a natureza permanente e transparente da *blockchain* pode levantar questões de privacidade.

Também existem preocupações legais e regulatórias que ainda precisam ser resolvidas (TAPSCOTT, TAPSCOTT, 2016), como o fato de não ser necessário uma autenticação prévia, onde seriam coletados dados pessoais, para usar as redes *blockchain*, como *Bitcoin* e *Ethereum*. Com isso é possível utilizar essas redes para cometer crimes financeiros, como lavagem de dinheiro e venda de produtos ilegais.

2.4. SMART CONTRACTS

Os contratos inteligentes, ou *smart contracts*, são protocolos computacionais que facilitam, verificam e aplicam a negociação e execução de um contrato, possibilitando transações confiáveis sem a necessidade de um intermediário (SZABO, 1996). Esses contratos “autônomos”, que possuem a habilidade de serem executados após uma transação de gatilho, com as regras definidas no acordo e especificadas em código, têm o potencial de aumentar a eficiência e reduzir a dependência em sistemas transacionais para cumprimento do acordo.

Para o desenvolvimento de contratos inteligentes, se faz necessário a utilização de uma plataforma que suporte-os. Atualmente uma das principais plataformas é a *Ethereum*, que introduziu um ambiente computacional completo e *turing-complete*, um sistema de regras de controle de dados que pode ser usado para simular qualquer máquina de Turing, na *blockchain*, chamado de *Ethereum Virtual Machine* (EVM). Em contraste com a *Bitcoin*, que possui uma linguagem de script limitada, a *Ethereum* permite a criação e execução de contratos inteligentes complexos (BUTERIN, 2014).

Smart contracts apresentam uma ampla gama de potenciais aplicações, como finanças, cadeia de suprimentos e governança. Eles podem ser usados para criar aplicações descentralizadas (dApps), organizações autônomas descentralizadas

(DAOs) e *tokens* não fungíveis (NFTs), entre outras aplicações (TAPSCOTT, TAPSCOTT, 2016).

Porém, questões legais e regulatórias em torno dos contratos inteligentes ainda precisam ser completamente resolvidas (HARROP, KNOTTENBELT, 2018). Essas organizações autônomas descentralizadas muitas vezes operam de forma independente de nenhum Estado, o que dificulta qualquer procedimento judiciário, elevando o nível de risco ao utilizar essas soluções. Por não necessitarem de uma estrutura organizacional para administração, contratos com erros ou vulnerabilidades podem apresentar perdas irreparáveis aos detentores dos mesmo, ainda mais agravado por conta de sua imutabilidade, fazendo com que a atualização de um contrato falho não seja possível, apenas a criação de um contrato substituto, invalidando o passado, corrigindo um possível erro.

2.5. HYPERLEDGER FABRIC

Hyperledger Fabric é uma plataforma de *blockchain* privada, destinada a aplicações empresariais, desenvolvida pela IBM e *Digital Asset* em 2015 e lançada pela *Linux Foundation* em 2017. Destaca-se de outras plataformas de *blockchain* por seu desenho modular e configurável, permitindo a implementação de um alto grau de confidencialidade, resiliência, flexibilidade e escalabilidade (ANDROULAKI et al., 2018). Além disso, essa plataforma disponibiliza uma visão atual da *blockchain*, chamada de *world state*, que mitiga a necessidade de percorrer toda a cadeia de transações para encontrar a versão mais atual dos dados.

O *Fabric* tem como uma das principais diferenças a outras plataformas de *blockchain* o seu suporte a contratos inteligentes escritos em linguagens de programação de propósito geral, como *JavaScript* e *Go*, e pelo uso de um mecanismo de consenso *pluggable*, ao contrário do *Ethereum* que suporta apenas linguagens mais específicas, como *Solidity*. Esta característica permite que diferentes mecanismos de consenso sejam usados dependendo das necessidades específicas da rede, sem a necessidade de um consenso de prova de trabalho intensivo em energia (CACHIN, 2016).

Uma das características principais do *HyperLedger Fabric* é a sua abordagem baseada em canais para a privacidade. Comparada a outras *blockchains*, que todas as transações são visíveis para todos os participantes, o *Fabric* permite a criação de

canais privados, em que apenas participantes autorizados podem ver as transações. Isso é particularmente útil para casos de uso em que a privacidade dos dados é essencial (RATHEE et al., 2020).

No entanto, a configuração e manutenção de uma rede *Fabric* pode ser complexa devido à sua natureza modular, que introduz um custo maior em sua implementação. Outro desafio é a realização de atualizações ou mudanças na rede, pois isso exige uma coordenação entre todos os membros, que pode ser demorado em uma rede grande (ANDROULAKI et al., 2018).

2.6. BIG DATA

Big data refere-se ao grande volume de dados, estruturados e não estruturados, que inundam os negócios diariamente. Estes dados são caracterizados por três principais atributos, conhecidos como os 3Vs: volume, velocidade e variedade (LANEY, 2001). Com essas 3 características principais, *big data* requer um processamento diferente do tradicional, necessitando, muitas vezes, da utilização do paradigma da computação distribuída para processar esses dados de forma a extrair valor dos mesmos.

Além da definição formal, também é possível classificar uma aplicação como *big data* pelo uso de ferramentas *NoSQL*, que representa soluções que não seguem o paradigma *SQL* tradicional, e pela utilização do paradigma *online analytical processing* (OLAP), ao invés do *online transaction processing* (OLTP), que é o sistema transacional mais comum, usado em plataformas *web* e em sistemas do dia-a-dia. Com isso podemos classificar nossa base de dados como *big data*, pois ela fará uso de ferramentas *NoSQL* e utilizará o sistema OLAP.

Com a crescente digitalização dos negócios e da sociedade, cada vez mais dados estão sendo criados e armazenados. Estes dados, quando adequadamente analisados, podem fornecer *insights* valiosos e orientar decisões estratégicas. Através do uso de técnicas de mineração de dados e aprendizado de máquina, padrões e conexões podem ser descobertos em conjuntos de *big data* que podem fornecer vantagem competitiva e otimização de processos (CHEN, MAO, LIU, 2014).

O uso do *big data* tem aplicações em uma variedade de campos. No comércio, por exemplo, o *big data* pode ser usado para análise preditiva, otimizando as operações e permitindo uma melhor segmentação de mercado. Na saúde, o *big*

data pode ser usado para melhorar os cuidados ao paciente, prever epidemias e realizar pesquisas médicas avançadas (RAGHUPATHI & RAGHUPATHI, 2014).

Apesar de seu potencial, o uso do *big data* também apresenta desafios. Questões de privacidade e segurança são preocupações significativas, pois grandes volumes de dados pessoais são frequentemente coletados e analisados. Além disso, a integridade e a qualidade dos dados podem afetar a eficácia dos *insights* gerados. Assim, o gerenciamento eficaz de *big data* requer uma combinação de tecnologia, governança de dados e habilidades analíticas (KITCHIN, 2014).

2.7. METABASE

Metabase é uma ferramenta de inteligência de negócios de código aberto que facilita o acesso e a análise de informações em uma organização, servindo como plataforma para análise de *big data*. Ela permite que usuários sem conhecimento técnico em SQL realizem consultas e obtenham insights a partir de seus dados. A interface intuitiva do *Metabase* e sua capacidade de integração com várias fontes de dados tornam-no uma escolha popular para empresas que buscam democratizar o acesso a informações analíticas.

Uma das principais características do *Metabase* é sua interface de usuário amigável, que permite a criação de *dashboards* e relatórios personalizados. Os usuários podem realizar consultas complexas com poucos cliques e visualizar os dados em diferentes formatos, como gráficos, tabelas e mapas. Além disso, o *Metabase* oferece recursos de filtragem e segmentação de dados, facilitando a análise específica de subconjuntos de dados.

No contexto empresarial, o *Metabase* é utilizado para monitorar KPIs (*Key Performance Indicators*), realizar análises de tendências de mercado, e entender melhor o comportamento do cliente. Ele também é útil em cenários de análise de dados operacionais, como monitoramento de desempenho de vendas, análise de eficiência de processos internos e otimização de cadeias de suprimentos.

Embora o *Metabase* seja uma ferramenta poderosa, ele apresenta desafios, como a necessidade de garantir a segurança dos dados e a gestão adequada de permissões de usuários. Além disso, com o avanço da tecnologia de dados, espera-se que o *Metabase* continue evoluindo, integrando mais recursos de

inteligência artificial e aprendizado de máquina para fornecer análises preditivas e insights mais profundos.

3. METODOLOGIA

A presente pesquisa tem como proposta principal disponibilizar uma plataforma de consultas direcionada aos gestores políticos, como o Ministro da Educação, para que seja possível visualizar dados educacionais armazenados na *blockchain* da Jornada do Estudante. Essa plataforma idealizada se dará com uma conexão dos dados da *blockchain* ao *Metabase*.

Atualmente o MEC já utiliza o *PowerBI* em alguns painéis de observabilidade, contudo o *Metabase* é uma ferramenta que apresenta uma vantagem por ser gratuito e, por causa de sua natureza de código aberto, possuir uma maior elasticidade para se conectar com os dados da *blockchain* construída sobre o *Hyperledger Fabric*.

A presente pesquisa deve se integrar ao projeto Jornada do Estudante, que é mantido por uma parceria realizada entre o Laboratório de Segurança em Computação (LabSEC), o Ministério da Educação (MEC), e Laboratório Bridge, e, de acordo com o MEC, mapeia todo o ciclo de vida da experiência educacional de um estudante, incluindo instituições educacionais frequentadas, desde a admissão até a formatura, seu desempenho e outras métricas, com uma plataforma robusta de análise de dados. O projeto Jornada do Estudante não engloba o acesso aos dados por uma ferramenta analítica, assim, existe a oportunidade da aplicação dos resultados dessa pesquisa para a integração da *blockchain* utilizada com uma plataforma de *BI* através do consumo dos dados mais recentes da *blockchain* e seu armazenamento em uma plataforma *NoSQL*.

Outra questão relevante a ser analisada na implantação de uma plataforma de consulta é a gestão educacional, que é um campo de prática e estudo que se preocupa com a organização, administração, planejamento e avaliação das instituições de ensino, incluindo a liderança das pessoas e recursos para alcançar os objetivos educacionais desejados (BUSH, 2003). Com essa ferramenta que visa integrar dados educacionais por meio de uma integração de *blockchain* com *big data*, o gestor terá um maior controle na gestão educacional, que proporciona uma tomada de decisão mais concisa e com bases analíticas sólidas e extraíndo um maior valor dos dados da *blockchain*.

4. DESENVOLVIMENTO

O Ministro da Educação é uma autoridade de extrema relevância ao se tratar de tomada de decisão para a área educacional, especialmente na área de planejamento acadêmico para o futuro, surgindo a necessidade de informações completas sobre a caminhada acadêmica dos alunos. Para tanto, a visão do progresso educacional pelo Ministro da Educação é custosa nos dias atuais, sendo realizada anualmente através do Censo da Educação Superior, pois é necessário coletar os dados em cada universidade a fim de centralizá-los no Inep. Cada universidade armazena seus dados de progresso educacional de forma independente, isso acarreta em um trabalho extra de normalização, padronização e condensação dos dados. Com isso, a velocidade das tomadas de decisão é executada de forma lenta.

A disponibilização de um ambiente de dados normalizado, padronizado e condensado, facilita o processo de tomada de decisões pelo Ministro da Educação, visto que essa forma de dados possui uma estrutura mais limpa e acessível. A *blockchain* busca trazer esse ambiente, exigindo padrões para inserções de dados por parte do participante, garantindo a estrutura dos dados. Além disso, é possível acessar todos os dados simultaneamente, visto que todas as inserções de todos os participantes estarão disponíveis na *blockchain*, porém, no caso da *blockchain* da Jornada do Estudante, esses dados estão em conformidade com a LGPD e não podem ser acessados a nível individual, apenas através de junções ou anonimização dos dados.

Com essa junção dos dados das universidades na *blockchain* da Jornada do Estudante o processo de análise de dados e tomada de decisão é facilitado devido ao novo acesso às informações, o que faz com que o Ministro da Educação tenha a sua disposição todos os dados a todo momento, favorecendo tomadas de decisões críticas e sensíveis. Esse novo formato de visualização, permite uma maior transparência nas informações, trazendo uma maneira de analisar os dados, de forma visual e já modelada, aliado com maior velocidade de atualização dos dados, os disponibilizando assim que são incluídos no *world state* da *blockchain*.

A representação visual dos dados por meio de *dashboards*, permite com que sejam identificados padrões nos dados, auxiliando no processo de correção de

anomalias e prevenindo erros futuros. Além de possibilitar uma monitoração constante do desempenho acadêmico dos alunos, possibilitando uma previsão de algum novo problema e, com base na centralização de informação, acaba por facilitar ações corretivas pelo Ministério da Educação e o Ministro da Educação.

Fazendo o uso de tecnologias de análise de dados, é possível maximizar a ingestão de dados e minimizar o tempo para disponibilidade de um dado. Buscando a utilização de ferramentas abertas e gratuitas, é possível recorrer ao *Metabase*, devido a sua ampla comunidade, seu acesso gratuito e por ser uma plataforma de código aberto, para consultar os dados da *blockchain*. É possível construir painéis com informações completas para serem analisadas pelos responsáveis, tornando o acesso à informação mais democrático dentro do MEC. O *Metabase* disponibiliza uma ampla configuração de permissões, que possibilita que sejam criados vários níveis de acesso, protegendo informações sensíveis de quem não deve ter acesso às mesmas. Assim, é possível realizar ajustes, de forma que o Ministro da Educação tenha o maior nível de acesso, podendo ver todas as informações disponíveis.

A partir da conexão da ferramenta analítica aos dados da *blockchain* privada, através da utilização do *Metabase* e do *world state* do *Hyperledger Fabric*, é possível garantir que a informação é válida e verificada por cada membro da rede, aumentando a confiabilidade dos dados e impedindo que informações não verdadeiras, ou adicionadas de forma não convencional, impacte a avaliação do progresso dos alunos.

4.1. RESULTADOS ESPERADOS

Com a implementação da abordagem exposta na pesquisa, espera-se uma otimização no processo de tomada de decisões em relação aos cursos superiores no Brasil, buscando uma realocação de recursos mais eficientes entre as universidades, tornando mais efetivos os programas de incentivo à educação.

Além disso, é esperado que exista uma nova visão em relação a formação de novos profissionais, possibilitando o adiantamento da formatura de alunos em cursos requisitados, como médicos durante uma crise sanitária ou professores de ensino básico para auxiliar em creches, permitindo uma alocação eficiente e eficaz de novos profissionais no mercado.

Outro propósito dessa ferramenta é fornecer uma visão ampla do desempenho dos alunos, permitindo que ações direcionadas sejam tomadas, reduzindo a evasão dos cursos e aumentando a satisfação dos alunos para com a universidade.

Em suma, observa-se que a proposta de combinar *big data* e *blockchain* para analisar dados educacionais surge da necessidade de gerenciar eficientemente dados educacionais descentralizados e padronizá-los. Esta abordagem oferece um meio mais preciso e conciso de tomar decisões baseadas em dados, fornecendo uma visão analítica abrangente dos dados educacionais.

Além disso, a utilização de painéis (*dashboards*) é uma parte crucial da proposta, pois eles fornecem uma interface visual para representar e analisar os dados coletados. Os painéis facilitam o acesso a informações complexas de uma maneira mais compreensível, permitindo que os gestores educacionais tomem decisões informadas com base em análises de dados detalhadas e atualizadas. Com isso, esses painéis ajudam a identificar tendências, padrões e áreas de melhoria no sistema educacional.

Essa proposta é valiosa por oferecer uma abordagem mais eficiente e baseada em dados para a tomada de decisões, melhorando a gestão educacional. A tecnologia *blockchain* fornece segurança e transparência desses dados, enquanto a centralização de informações anteriormente dispersas facilita o acesso e a análise. Além disso, ao fornecer *insights* detalhados sobre métricas educacionais, como o desempenho dos alunos, esta abordagem tem o potencial de melhorar significativamente o sistema educacional.

4.2. IMPLEMENTAÇÃO

A partir dos dados armazenados na *blockchain*, foi realizada uma extração do *world state*, que representa o estado mais atualizado da *blockchain*, recuperando os dados mais recentes de cada contrato. Com isso, esses dados foram escritos no *MongoDB*, uma ferramenta de banco de dados *open-source* gratuita que possui um formato de armazenamento similar ao *world state* do *Hyperledger Fabric*, denominado *json document*, que consiste em um documento “chave: valor”, onde cada documento representa um registro no banco de dados.

Após a extração e armazenamento dos dados, foi utilizado o *Metabase* para analisar as novas tabelas e extrair informação dos dados armazenados. O *Metabase* possui diversos *drivers* para conexão com diversas bases de dados, porém não fornece suporte aos dados da *blockchain* da Jornada do Estudante. Com isso, foi necessário utilizar uma ferramenta para extração de dados diretamente da *API (Application Programming Interface)*, que é um conjunto de rotinas e padrões que permitem que dois ou mais aplicativos se comuniquem entre si, da Jornada do Estudante e escrever esses dados no *MongoDB*.

4.2.1. Público alvo

A ferramenta é destinada para as autoridades responsáveis pela tomada de decisão dentro do Ministério da Educação, com foco no Ministro da Educação, e permite que vários servidores do Ministério tenham acesso aos painéis com permissões de visualizações em diversos níveis, protegendo dados sensíveis e especificando a função de cada um dentro do sistema. Os painéis têm como foco o Ministro da Educação pois ele é quem detém o poder de decisão dentro do MEC, decidindo quais ações serão tomadas para auxílio dos alunos.

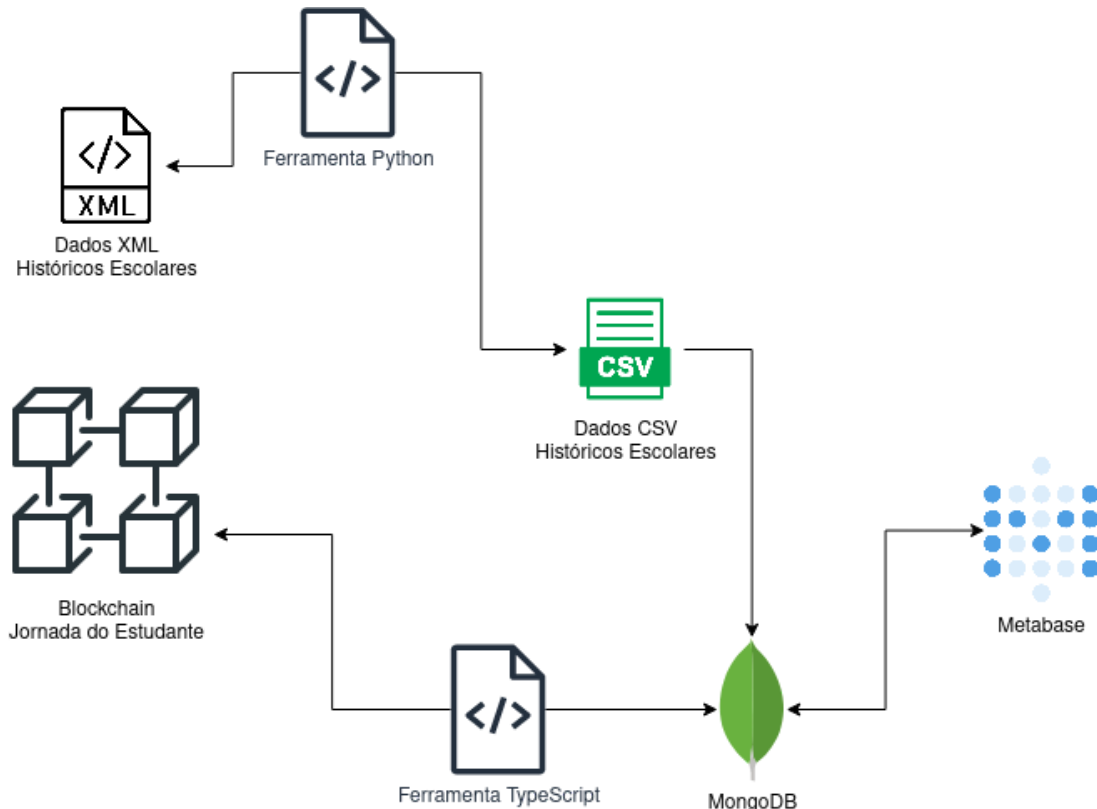
Esses painéis também possuem valor para os reitores das universidades e os diretores de centro. Essa visão detalhada do progresso dos alunos possibilita que programas de incentivo sejam executados com maior precisão e fornece uma plataforma onde é possível medir o impacto dessas medidas semestre-a-semester, aumentando ainda mais a efetividade dessas intervenções.

4.2.2. Protótipo

Para realização do protótipo, foi utilizada a rede de testes da Jornada do Estudante do MEC, disponibilizada pelo Laboratório de Segurança em Computação (LabSEC) da UFSC e preenchida com dados fictícios, com o intuito de preservar dados sensíveis. A *blockchain* de teste utilizou o *Minifabric*, que fornece um ambiente do *Hyperledger Fabric* de forma compacta e versátil, utilizando *Docker*, uma plataforma de software livre que permite aos desenvolvedores criar, implementar e executar aplicativos em contêineres, que é um pacote leve e empacotado de software que inclui tudo o que o aplicativo precisa para funcionar,

incluindo código, bibliotecas e configurações, para fornecer uma solução prática para instalação local, sem necessidade de um ambiente com vastos recursos de processamento e memória. Os dados utilizados foram construídos de forma a replicar dados de produção, porém sem informações reais.

Figura 4 - Fluxograma de ingestão dos dados.



Fonte: De autoria própria.

Como demonstrado na Figura 4, os dados da *blockchain* foram extraídos utilizando uma ferramenta disponibilizada pelo LabSEC, escrita em *TypeScript*, um superconjunto sintático de *JavaScript*, que fornece uma tipagem adicional à linguagem. Além disso, foi desenvolvida uma ferramenta escrita em *Python* (disponível em <https://github.com/pedronobrega/ufsc-tcc-pedro-nobrega>), para realizar a extração dos históricos escolares, armazenados em *XML* (*eXtensible Markup Language*), uma linguagem de marcação que pode ser usada para representar dados de uma forma estruturada, e realizar a conversão desses dados para *CSV* (*Comma-separated values*), um formato de arquivos onde os valores são separados por vírgula. Esses dados foram armazenados no *MongoDB*, pois é o

banco com maior semelhança ao formato de dados do *world state* do *Hyperledger Fabric*.

Utilizando o Metabase, foi possível acessar os dados armazenados no *MongoDB*. Com esses dados foi necessário fazer uma modelagem do banco de dados, de forma a desaninhar os dados aninhados, que são dados organizados em níveis, onde cada nível contém dados relacionados ao nível anterior, em uma única tabela que contém apenas um nível, de forma que o *Metabase* consiga acessar os dados de forma direta, sem precisar processar os dados ou construir consultas complexas para planificar os dados. Essa modelagem possibilitou a construção de painéis para visualização dos dados.

Para essa pesquisa, foram escolhidos os seguintes painéis, dando foco para informações referentes aos alunos:

- Número de alunos por sexo e idade;
- Nota média por disciplina;
- Desempenho por região de origem do aluno;
- Nota média por disciplina por período.

Todos esses painéis fazem uso da tabela referente ao histórico escolar do aluno. Além destas visualizações, é possível filtrar essas informações nos seguintes níveis:

- UF (Unidade Federativa);
- IES (Instituição de Ensino Superior);
- Curso;
- Disciplinas;
- Período letivo;
- Sexo.

Com essa granularidade, é possível investigar como está o progresso dos alunos, observando, por exemplo, uma alta taxa de reprovação em certa disciplina, ao comparar o desempenho médio entre os períodos. Outra possibilidade é realizar uma análise demográfica dos alunos, comparando a nota média nas disciplinas em relação à UF de origem do aluno, possibilitando, por exemplo, uma visualização do impacto de estudar em uma universidade tão distante de sua família.

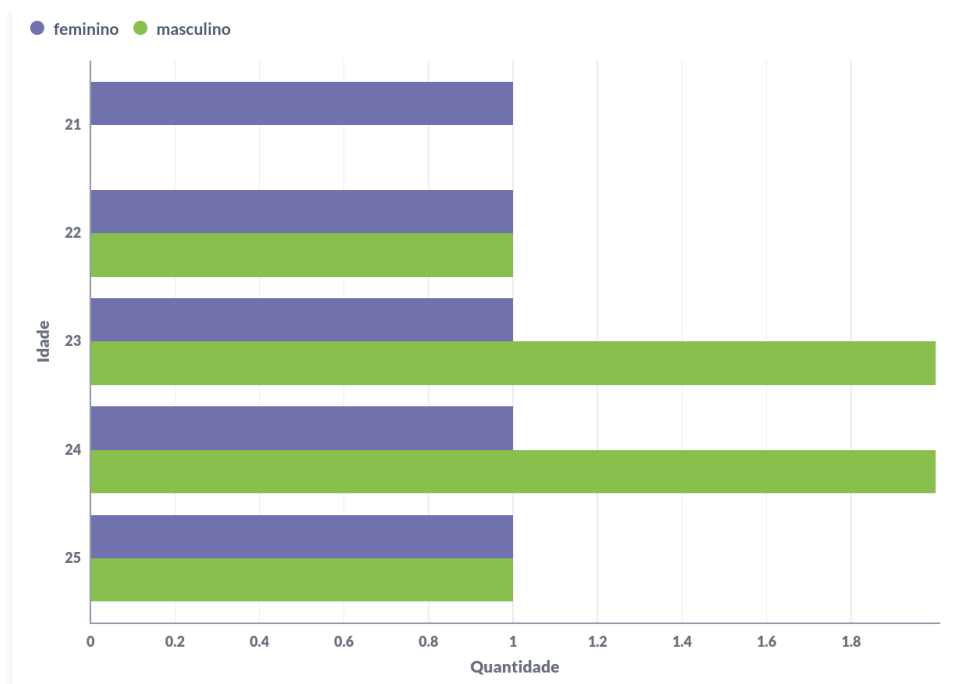
Com as visões de dados atualizados em formato de gráficos é possível analisar de forma profunda como está o desempenho dos alunos nos cursos, sendo transparente em vários fatores, como o desempenho de alunas do sexo feminino em

cursos que apresentam uma maioria masculina e vice-versa, assim como é possível visualizar o impacto de reprovações em disciplinas importantes no curso, como disciplinas que são pré-requisito de várias outras no decorrer do currículo, e entender como mitigar possíveis consequências dessas reprovações. Com base nessas informações, é possível realizar filtros específicos, observando estas estatísticas em um nível ainda mais granular, selecionando os dados por sexo do aluno ou pelo seu período de ingresso, auxiliando a visão completa de como certos grupos estão desempenhando sua graduação.

5. RESULTADOS E DISCUSSÃO

Com base na implementação dos painéis e na utilização de dados de exemplo, seguindo o formato dos dados reais da *blockchain*, foram obtidos os diversos gráficos para demonstrar os dados de uma forma clara e concisa.

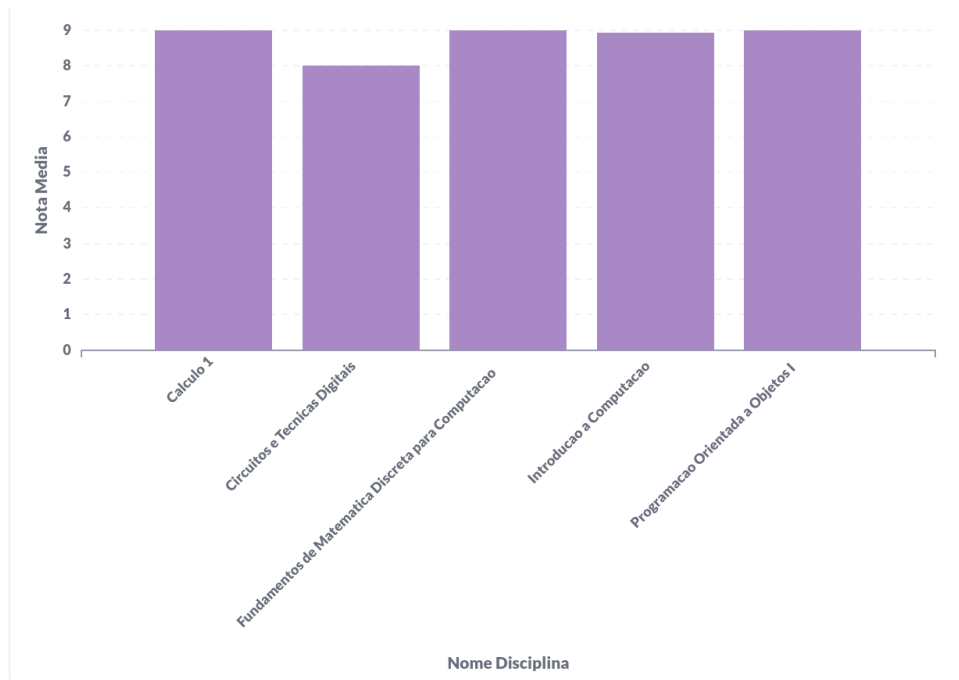
Figura 5 - Alunos por sexo e idade.



Fonte: De autoria própria.

No gráfico apresentado na Figura 5 é possível ter uma visão da distribuição demográfica dos alunos, para visualizar a quantidade de alunos por idade e sexo. Esses dados podem ser filtrados por alguns atributos desses alunos, como o curso, a disciplina e o período letivo que a disciplina foi cursada.

Com base no gráfico da Figura 5, é possível verificar como está a distribuição dentro de cada universidade, fazendo com que seja possível direcionar esforços para fornecer oportunidades iguais para alunos de qualquer idade e qualquer sexo, fornecendo oportunidades iguais para todos.

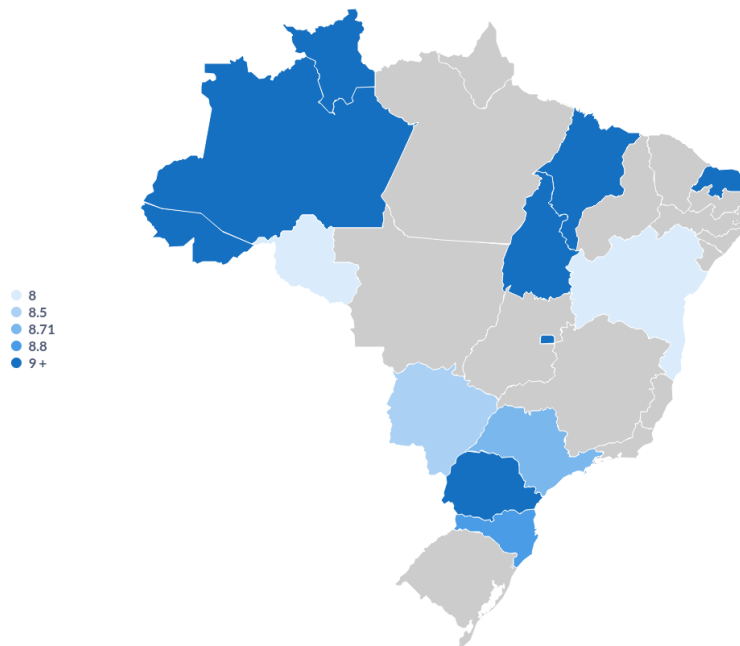
Figura 6 - Nota média por disciplina.

Fonte: De autoria própria.

Com base na visualização do gráfico apresentado na Figura 6, podemos observar a média das notas dos alunos de cada disciplina, verificando o desempenho dos mesmos nas disciplinas cursadas. Como demonstra a Figura 6, o gráfico permite o filtro pelo curso, período letivo e o sexo dos alunos, assim é possível compreender a performance dos estudantes conforme avançam no curso.

Por meio de gráficos com uma visibilidade granular, é possível identificar pontos de inflexão no curso, ocasionando em desistências. Assim, pode-se direcionar esforços para corrigir como as disciplinas são ministradas e/ou avaliadas, aumentando a taxa de permanência no curso.

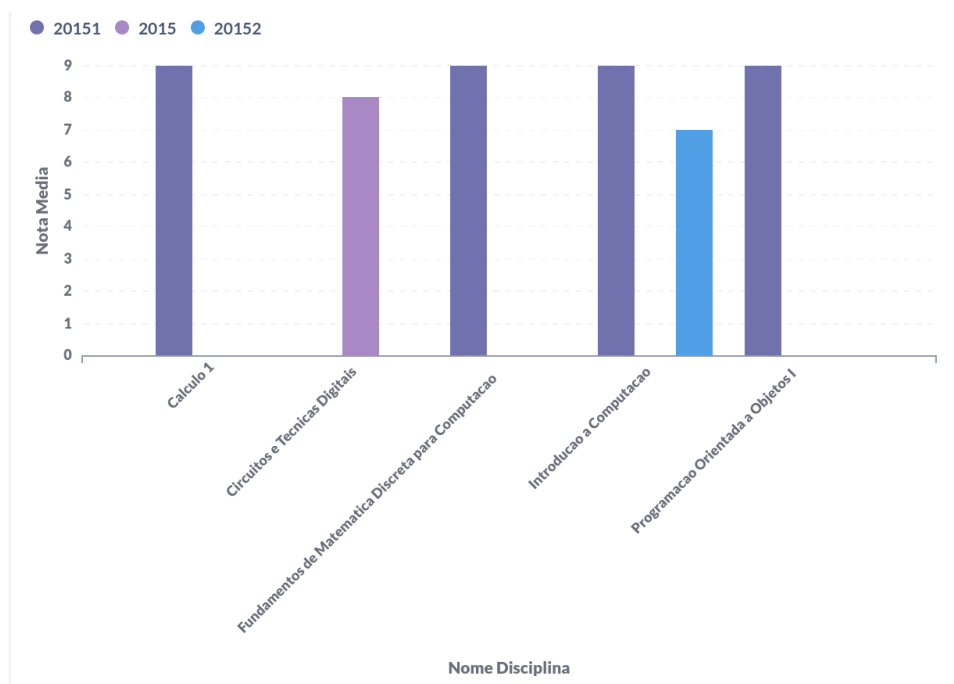
Figura 7 - Desempenho por estado de origem.



Fonte: De autoria própria.

O gráfico exibido na Figura 7 apresenta uma visualização utilizando mapa de calor e demonstrando a nota média pelo estado de origem de cada estudante, podendo ser filtrado por curso, disciplina, período letivo e sexo do aluno, esse gráfico demonstra o desempenho acadêmico dos alunos de forma granular, podendo observar como cada grupo se desempenha no decorrer do percurso acadêmico.

A visualização utilizando um mapa de calor é útil para compreender como auxiliar alunos que mudam sua residência com o intuito de cursar uma universidade em outro estado e acabam por romper sua rede de apoio familiar, o que pode ocasionar em uma diminuição das notas e um acréscimo na taxa de evasão do curso.

Figura 8 - Nota média das disciplinas por período.

Fonte: De autoria própria.

A visualização apresentada no gráfico da Figura 8 fornece um entendimento de como cada disciplina está sendo aproveitada pelos alunos ao longo do tempo. Podendo ser filtrada por sexo do aluno e por períodos letivos específicos, é possível acompanhar a efetividade das metodologias aplicadas em cada disciplina.

Com o gráfico da Figura 8 é possível observar possíveis dificuldades na passagem de conhecimento das disciplinas, acompanhar a mudança de metodologias e/ou professores responsáveis. Assim, é possível planejar como as disciplinas são ministradas, com o intuito de aumentar o aproveitamento estudantil e a taxa de permanência nos cursos.

Com base nos resultados obtidos, é possível concluir que esses gráficos apresentam uma visão completa do que pode ser feito com os dados da *blockchain* do Jornada do Estudante, fornecendo informações valiosas de como os alunos estão aproveitando a universidade e os cursos ofertados.

Podemos analisar também o papel do *Metabase* na construção dos gráficos. Essa ferramenta possui um controle de acesso à cada base de dados, podendo ser configurada quem tem acesso a cada um dos gráficos e das *dashboards*, assim preservando a privacidade dos estudantes e disponibilizando para os funcionários do MEC apenas os gráficos relevantes para sua função. Porém, alguns pontos devem

ser levantados, como a falta de possibilidade de extração dos gráficos, que permitiria fazer backup externo dos mesmos, assim como a falta de uma forma direta de consulta à *json documents*, o formato adotado pelo *world state* do *Hyperledger Fabric* e do *MongoDB*, na qual não é possível consultar os dados aninhados, sendo necessário um trabalho de planificação dos dados e a criação de novas tabelas para consulta dos mesmos. Outro ponto importante é a falta de possibilidade de criação de gráficos complexos com maior dinamismo com relação aos filtros, que possibilitam uma visualização mais completa e com maior valor agregado.

5.1. CONSIDERAÇÕES FINAIS

Ao longo desta pesquisa, foi possível observar a importância da utilização de tecnologias como *big data* e *blockchain* na análise de dados educacionais. Por meio da análise de dados de teste, já padronizados, do Jornada do Estudante, foi possível identificar pontos de inflexão no curso, ocasionando em desistências, e direcionar esforços para corrigir como as disciplinas são ministradas e/ou avaliadas, aumentando a taxa de permanência no curso.

Além disso, a utilização da tecnologia *blockchain*, aliada ao controle de acessos do *Metabase*, permitiu a segurança e transparência dos dados educacionais, garantindo que apenas pessoas autorizadas tenham acesso a essas informações. Isso é especialmente importante em um contexto em que a privacidade e a segurança dos dados são cada vez mais valorizadas.

Por fim, é importante destacar que os resultados obtidos nesta pesquisa podem ser utilizados para melhorar a qualidade da educação no Brasil. Ao compreender a performance dos estudantes conforme avançam no curso, é possível identificar as disciplinas que apresentam maiores dificuldades e direcionar esforços para melhorar o ensino dessas matérias. Dessa forma, espera-se contribuir para a formação de profissionais mais capacitados e preparados para enfrentar os desafios do mercado de trabalho.

6. CONCLUSÃO

Em suma, a presente pesquisa apresentou um sistema inovador para análise de dados educacionais utilizando tecnologias como o *Metabase* e a *blockchain* do projeto Jornada do Estudante. Através da utilização dessas tecnologias, foi possível garantir a segurança e consistência dos dados, bem como identificar pontos de inflexão nos cursos que podem ser corrigidos para melhorar a qualidade da educação.

Ao conectar o *Metabase* à base de dados da *blockchain*, foi possível modelar os dados a fim de construir novas tabelas que satisfizeram a necessidade de consulta de dados, fornecendo painéis de visualizações complexos e com uma variedade de informação. Com base nesses painéis, é possível ter uma visão do que se pode alcançar com esses dados e como os profissionais da educação podem utilizar essas informações no processo de tomada de decisão dentro do Ministério da Educação.

Os resultados obtidos neste trabalho são promissores e podem ser utilizados para melhorar a qualidade da educação no Brasil. Ao compreender a performance dos estudantes e identificar as disciplinas que apresentam maiores dificuldades, é possível direcionar esforços para melhorar o ensino dessas matérias e, assim, formar profissionais mais capacitados e preparados para enfrentar os desafios do mercado de trabalho.

6.1 TRABALHOS FUTUROS

Diante dos resultados apresentados na pesquisa, surge a necessidade de continuação do desenvolvimento com projetos na mesma área. Uma oportunidade de trabalho futuro a nível de mestrado é a conexão direta da *blockchain* com o *Metabase*, criando um novo conector a fim de lidar com todas as chamadas necessárias para visualizar os dados de forma ágil e completa.

Outro possível trabalho futuro que trará um valor agregado para a solução proposta, é um projeto que busca adicionar observabilidade à *blockchain* e à

plataforma analítica, extraindo informações do uso no dia-a-dia, assim como garantindo a integridade e segurança dos dados.

Por fim, é importante destacar que este trabalho é apenas o começo de uma jornada que pode levar a avanços significativos na área de educação. A utilização de tecnologias como *big data* e *blockchain* pode revolucionar a forma como os dados educacionais são coletados e analisados, permitindo uma compreensão mais profunda do processo de ensino e aprendizagem. Esperamos que este trabalho possa inspirar outros pesquisadores a explorar ainda mais essas tecnologias e contribuir para a melhoria da educação no Brasil e no mundo.

REFERÊNCIAS

ANDROULAKI, E. et al. **Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains**. In: Proceedings of the Thirteenth EuroSys Conference (EuroSys '18). New York: Association for Computing Machinery, Article 30, p. 1–15, 2018.

ANTONOPOULOS, A. M. **Mastering Bitcoin: Unlocking Digital Cryptocurrencies**. Sebastopol: O'Reilly Media, 2014.

BUSH, T. **Theories of Educational Leadership and Management**. London: Sage, 2003.

BUTERIN, V. **Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform**. Ethereum White Paper, 2014.

CACHIN, C. **Architecture of the Hyperledger Blockchain Fabric**. In: Workshop on Distributed Cryptocurrencies and Consensus Ledgers, 2016.

CHEN, M.; MAO, S.; LIU, Y. **Big Data: A Survey**. Mobile Networks and Applications, v. 19, n. 2, p. 171–209, 2014.

COULOURIS, G.; DOLLIMORE, J.; KINDBERG, T. **Distributed Systems: Concepts and Design**. London: Pearson Education, 2005.

Docker. **Docker**. Acesso em: 15/12/2023. Disponível em: <https://docker.com>.

EMMERICH, W. **Software engineering and middleware: a roadmap**. In: Proceedings of the Conference on the Future of Software Engineering, Limerick, Ireland, 2000.

HARROP, W.; KNOTTENBELT, W. **The DAO Attack: Mathematical Modelling and Lessons Learned**. In: Gervais A., Livshits B. (eds) Financial Cryptography and Data Security. FC 2018. Lecture Notes in Computer Science, vol 10958. Cham: Springer, 2018.

HyperLedger Fabric. **HyperLedger Fabric**. Acesso em 15/12/2023. Disponível em: <https://hyperledger.org/projects/fabric>.

HypeprLedger Labs. **Minifabric**. Acesso em: 15/12/2023. Disponível em: <https://github.com/hyperledger-labs/minifabric>.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Censo da Educação Superior**. Acesso em: 15/12/2023. Disponível em: <https://gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior>.

KITCHIN, R. **Big Data, new epistemologies and paradigm shifts**. Big Data & Society, v. 1, n. 1, 2014.

LANEY, D. **3D data management**: Controlling data volume, velocity, and variety. META Group Research Note, n. 6, 2001.

LÜCK, Heloísa. **A evolução da gestão educacional a partir de mudança paradigmática**. Gestão em Rede, [s.l.], n. 3, p. 13-18, nov. 1997.

LYNCH, N. **Distributed Algorithms**. San Francisco: Morgan Kaufmann Publishers Inc., 1996.

MENEZES, A.; VAN OORSCHOT, P.; VANSTONE, S. **Handbook of Applied Cryptography**. Boca Raton: CRC Press, 1996.

Metabase. **Metabase**. Acesso em 15/12/2023. Disponível em: <https://metabase.com>.

Microsoft. **PowerBI**. Acesso em 15/12/2023. Disponível em: <https://powerbi.microsoft.com/pt-br>.

Ministério da Educação. **Jornada do Estudante**. Acesso em 15/12/2023. Disponível em: <https://gov.br/mec/pt-br/jornadadoestudante>.

Ministério da Educação. **Portal Único de Acesso ao Ensino Superior**. Acesso em 15/12/2023. Disponível em: <https://accessunico.mec.gov.br/busca>.

NAKAMOTO, S. **Bitcoin: A Peer-to-Peer Electronic Cash System**, 2008.

NETO, N. N. **Criptografia em Comunicação de Dados**. São Paulo, SP: 2002.

RAGHUPATHI, W.; RAGHUPATHI, V. **Big data analytics in healthcare**: promise and potential. Health Information Science and Systems, v. 2, n. 1, p. 3, 2014.

RATHEE, G.; SHARMA, H.; SHARMA, V.; IYER, S. **Performance Benchmarking and Optimizing Hyperledger Fabric Blockchain Platform**. In: 2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID), 2020.

RESCORLA, E. SSL and TLS: **Designing and Building Secure Systems**. Boston: Addison-Wesley Professional, 2001.

Revisa Med. **Saiba onde estão as vagas de Residência Médica no Brasil**. Acesso em 15/12/2023. Disponível em: <https://revisamed.com.br/residencia-medica/vagas-de-residencia-medica>.

RIVEST, R.; SHAMIR, A.; ADLEMAN, L. **A Method for Obtaining Digital Signatures and Public-Key Cryptosystems**. Communications of the ACM, v. 21, n. 2, p. 120-126, 1978.

SCHNEIER, B. **Applied Cryptography: Protocols, Algorithms, and Source Code in C**. New York: Wiley, 1996.

SAGIROGLU, S.; SINANC, D. **Big data**: A review. In 2013 International Conference on Collaboration Technologies and Systems (CTS), p. 42-47, 2013.

Statista. **Size of the blockchain technology market worldwide in 2018 and 2019, with forecasts from 2020 to 2025**. Acesso em: 15/12/2023. Disponível em: <https://statista.com/statistics/647231/worldwide-blockchain-technology-market-size/>

SZABO, N. **Smart Contracts**: Building Blocks for Digital Markets. Extropy, n. 16, 1996.

TANENBAUM, A. S.; VAN STEEN, M. **Distributed Systems**: Principles and Paradigms. Upper Saddle River: Pearson Prentice Hall, 2007.

TAPSCOTT, D.; TAPSCOTT, A. **Blockchain Revolution**: How the Technology Behind Bitcoin is Changing Money, Business, and the World. New York: Portfolio, 2016.

YAGA, Dylan et al. **Blockchain technology overview**. arXiv preprint arXiv:1906.11078, 2019.

APÊNDICE A - ARTIGO DA MONOGRAFIA

Big data e Blockchain: um sistema para análise de dados educacionais.

Pedro Henrique Dias Nobrega

¹DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC)

Abstract. *With the decentralization of the educational system, distributed throughout the country, there is a need for a unified consultation of data at the national level. Due to the lack of data unity and the very nature of the information, there is a proposal to unify these data on a Blockchain, bringing standardization and strict control to the participants' data. The term Blockchain can be defined as a ledger implemented in a distributed way, usually without a central authority. Originally coined by Satoshi Nakamoto — the pseudonym of the supposed creator of Bitcoin — the Blockchain's function is to record data in a way that is resistant to time and potential attempts to modify it, using a consensus mechanism to record new blocks in the distributed ledger. With the data unified and available, it is possible to use a system to consult them through some tool, which supports a large amount of data, a subsequent increase of this information, and that can communicate with the used Blockchain. This difficulty in analysis makes it paramount to use Big Data, which is the term that defines massive data sets that have a great variety and complex structures with difficulties in ordering, analyzing, and visualizing for future processing or results. With a large amount of information available, there is a need to perform analyses, models, and explorations in these data sets. Thus, the field of Big Data emerges, focused on providing an efficient and effective way to manipulate and store these data. Using tools like Metabase, specifically developed for the analysis of large data, it is possible to have an analytical view of educational information. In this way, it is possible to ensure greater educational management over Brazilian education, which would allow for a more precise and concise decision-making system. With this, a system was created for the visualization of these data in the form of dashboards, presenting various panels, which can be used to improve the decision-making process within the Ministry of Education and by the Minister of Education.*

Resumo. *Com a descentralização do sistema educacional de ensino, distribuído em todo o país, observa-se a necessidade de uma consulta unificada dos dados a nível nacional. Por causa da carência de união dos dados e da própria natureza das informações, surge a proposta de unificar estes dados em uma Blockchain, trazendo padronização e controle rígido aos dados dos participantes. Pode-se definir o termo Blockchain como um livro-razão implementado de uma maneira distribuída, geralmente sem uma autoridade central. Tendo sido originalmente cunhado por Satoshi Nakamoto — pseudônimo do suposto criador do Bitcoin — a Blockchain tem como função registrar dados de forma resistente ao tempo e a possíveis tentativas de modificação dos mesmos, usando mecanismo de consenso para registrar novos blocos no livro-razão distribuído. Já com os dados unificados e disponíveis, é possível utilizar um sistema de consulta aos mesmos por meio de alguma ferramenta, que suporte uma grande quantidade de dados, um posterior incremento dessas informações e que consiga conversar com a Blockchain utilizada. Essa dificuldade de análise faz com que seja primordial o uso de Big Data, que é o termo que define conjuntos de dados massivos que possuem uma grande variedade e estruturas complexas com dificuldades de ordenamento, análise e visualização para futuros processamentos ou resultados. Com uma grande quantidade de informações disponíveis, considera-se a necessidade de fazer análises, modelos e explorações nesses conjuntos de dados. Com isso, surge a área de Big Data, focada em fornecer*

uma maneira eficiente e eficaz de manipular e armazenar esses dados. Utilizando ferramentas como o Metabase, desenvolvida especificamente para análise de grandes dados, é possível ter uma visão analítica das informações educacionais. Dessa forma, é possível garantir uma maior gestão educacional sobre o ensino brasileiro, que possibilitaria um sistema de decisão mais preciso e conciso. Com isso foi criado um sistema para visualização desses dados em formato de dashboards, apresentando painéis variados, que poderão ser utilizados para aperfeiçoar o processo de tomada de decisões dentro do Ministério da Educação e pelo Ministro da Educação.

1. INTRODUÇÃO

O Brasil possui um sistema educacional descentralizado, apresentando 2574 universidades espalhadas pelo país, fazendo com que existam diversos registros dispersos do desempenho acadêmico dos alunos, assim como dados demográficos e de permanência, nos cursos de graduação, que dificultam uma possível observação analítica sobre o sistema como um todo (INEP, 2021). Com o objetivo de tornar a abertura de processos regulatórios da educação superior no Brasil, em 2007 foi criado o e-MEC, visando dar agilidade e eficiência à comunicação entre o Ministério da Educação (MEC) e as instituições de ensino superior (IES). Com base na dificuldade enfrentada pelo MEC, no artigo “*Blockchain and smart contracts for higher education registry in Brazil*”, os autores propuseram a criação de um novo modelo que faz uso de tecnologias de *blockchain*, e da natureza do próprio sistema para criar um mecanismo de consenso e padronização dos dados. Podendo ser definido como um livro-razão digital implementado de uma maneira distribuída, geralmente sem uma autoridade central, a tecnologia *blockchain* com ferramentas de *big data* pode ser uma solução para o problema da informação dispersa (YAGA, MELL, ROBY E SCARFONE, 2019). Como solução ao problema dos dados despadronizados e sem coexistência, nasceu o Jornada do Estudante, que através de uma *blockchain* privada, uma *blockchain* onde os participantes podem ser escolhidos e as transações são visíveis apenas por membros credenciados, criou padrões para o armazenamento de dados e trouxe uma fonte de dados comum, sendo estes as próprias transações dentro da *blockchain*. Assim, o projeto Jornada do Estudante trouxe um novo meio de acompanhamento dos dados estudantis do Brasil, unificando-os em uma única plataforma que pode ser consultada por membros credenciados. O termo *big data* define conjuntos de dados massivos que possuem uma grande variedade e estruturas complexas com dificuldades de ordenamento, análise e visualização para futuros processamentos ou resultados destinados a facilitar que um usuário com conhecimentos de análise de dados consiga navegar pelo universo de dados diversos, incluindo dados provenientes de *blockchains*. Buscando mecanismos de decisão mais precisos e concisos, surge a necessidade de criar uma integração entre soluções de *big data* e *blockchain* possibilitando uma visão analítica dos dados. Ferramentas de *big data*, como por exemplo o *Metabase*, são uma maneira *open-source*, ou seja, de código aberto, de facilitar a análise de informações. A *blockchain* privada proposta faz uso do *Hyperledger Fabric*, que é uma plataforma para soluções baseadas em livro-razão distribuído, sustentada por uma arquitetura modular que entrega um alto grau de confidencialidade, resiliência, flexibilidade e escalabilidade e foi projetada para suportar implementações para diversos componentes. Pela natureza educacional da *blockchain* analisada durante esta pesquisa, foi constatado que a gestão educacional é uma etapa essencial na implementação do projeto de integração de *big data* e *blockchain* do Jornada do Estudante, pois está diretamente relacionada com a obtenção de dados educacionais de instituições de ensino, e pode ser definida como o reconhecimento da importância da participação consciente e esclarecida das pessoas nas decisões sobre a orientação e planejamento de seu trabalho (LÜCK, Heloísa). Diante do problema dos dados descentralizados, da falta de padrão dentre as universidades, e buscando fornecer uma melhor

gestão educacional, esta pesquisa busca solucionar o processo de conexão entre esses dois universos, propondo uma maneira de consultar os dados sequenciais padronizados da *blockchain* da Jornada do Estudante de forma eficiente e eficaz, trazendo *big data* e análise de dados ao mundo educacional.

1.1 JUSTIFICATIVA

Este projeto busca analisar dados educacionais, auxiliar na tomada de decisão, a partir de informações do desempenho dos alunos com base em diversos fatores, como o estado de origem do estudante e sua idade. Com isso, o processo de decisões internas de educação do Brasil será aperfeiçoado, com base em dados reais e coesos sobre o desenvolvimento acadêmico dos alunos da rede pública. Com essas informações em mãos, será possível moldar e estabelecer padrões acelerando a evolução do sistema educacional. De acordo com o *Statista*, 2023, a *blockchain* tem apresentado um crescimento exponencial, com um valor de mercado previsto de 40 bilhões de dólares em 2025, fazendo com que o tema tenha um desenvolvimento constante, surgindo novas soluções a todo momento. Com isso surge a necessidade de analisar os dados gerados para extrair valor desse meio e permitir aos servidores do Ministério da Educação (MEC) uma visão sistemática do progresso estudantil, fornecendo mais uma fonte de informação no processo de tomada de decisão. Conforme os dados da Demografia Médica no Brasil, 2020, “O Brasil deixou de formar, em cinco anos, quase 36 mil médicos especialistas, considerando as vagas autorizadas, mas não ocupadas.”. Essa taxa de vacância poderia ser mitigada fazendo uso da solução proposta neste projeto, onde seria possível analisar, por exemplo, a real quantidade de residentes que entrariam no processo seletivo e realocar as vagas conforme a necessidade. Essa lacuna de informação faz com que profissionais que desejam se especializar e desempenhar um papel importante na sociedade, não consigam vagas em suas áreas. Com a solução proposta e um cruzamento de dados com informações do Ministério da Saúde, seria possível prever essas oportunidades e preparar a oferta de vagas para suportar essa demanda.

1.2 OBJETIVOS

Para o desenvolvimento desta pesquisa foram definidos os seguintes objetivos: geral e específicos.

1.2.1 Objetivo geral

Esta pesquisa tem como objetivo geral a análise dos dados educacionais da rede *blockchain* da Jornada do Estudante através de uma interface de consulta de *big data*.

1.2.2 Objetivos específicos

- Conectar uma ferramenta analítica, que fornecerá a plataforma de análise de dados, à *blockchain*;
- Modelar os dados de uma *blockchain* a fim de satisfazer as necessidades específicas relacionadas à gestão educacional;
- Construir um *dashboard* com exemplos visuais de informações possíveis a serem obtidas com os dados da *blockchain*;

1.3 METODOLOGIA DE PESQUISA

A pesquisa possui abordagem qualitativa, em que a primeira etapa consistiu em leituras exploratórias do tema, buscando o que está sendo desenvolvido na área e identificando o estado da arte, extraindo os conceitos de *blockchain*, assim como obter o estado da arte de *big data* e ferramentas analíticas em processos de gestão de conhecimento. A partir disso, é feita

uma observação a fim de identificar como as tecnologias propostas na pesquisa estão sendo utilizadas pela comunidade, com aprofundamento em como adaptá-las para as necessidades da pesquisa. Com os resultados qualitativos, foi realizada uma análise quantitativa com objetivo de extrair métricas relacionadas às ferramentas selecionadas, de forma a fundamentar a escolha da mesma ou expandir o leque de opções para novas ferramentas, descartando as que possuem performance abaixo do esperado ou as que possuem custos acima do esperado para a solução proposta.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, o leitor será introduzido aos conceitos base para entendimento do funcionamento das tecnologias utilizadas no desenvolvimento do trabalho, assim como os motivos dessas escolhas.

2.1 CRIPTOGRAFIA ASSIMÉTRICA

O termo criptografia define um conjunto de técnicas que visam transformar uma mensagem de modo a torná-la incompreensível para quem não esteja autorizado a lê-la. Essa transformação é realizada por meio de um algoritmo, que é um conjunto de regras matemáticas que alteram a forma da mensagem (NETO, 2002). A Criptografia Assimétrica, também conhecida como criptografia de chave pública, é uma técnica de criptografia que utiliza um par de chaves (uma pública e uma privada) que são matematicamente relacionadas, porém não idênticas. Assim, tudo que a chave privada criptografa a pública descriptografa, e vice-versa. A chave pública é usada para criptografar mensagens, enquanto a chave privada é usada para descriptografá-las (SCHNEIER, 1996), fornecendo sigilo às informações criptografadas, assim como o inverso, utilizando a chave privada para criptografar mensagens e a pública para descriptografá-las, fornecendo autenticidade às informações. Essa forma de criptografia proporciona uma segurança pela dificuldade de deduzir a chave privada a partir da chave pública. A Figura 1 representa como funciona esse tipo de criptografia na prática, onde Bob utiliza a chave pública de Alice para criptografar a mensagem, que apenas pode ser descriptografada pela chave privada de Alice.

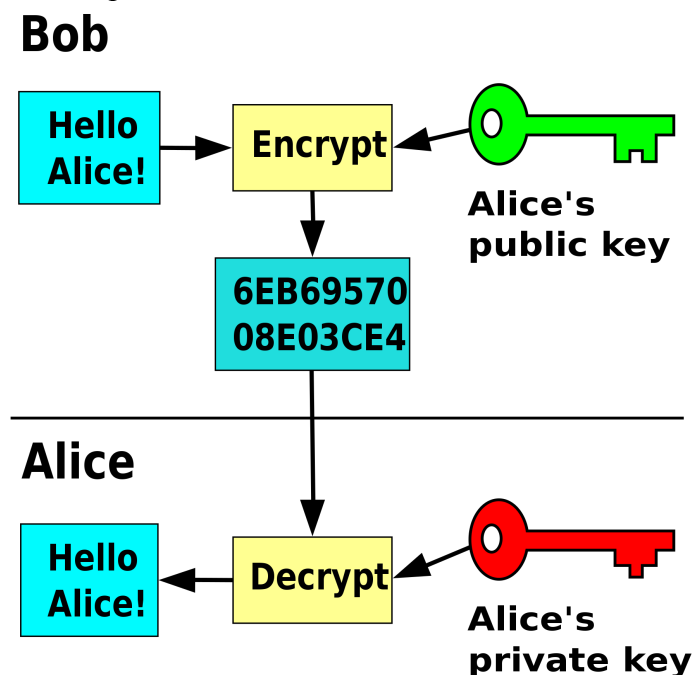


Figura 1 - Criptografia de Chave Pública.

Uma das formas de criptografia assimétrica amplamente utilizadas para proteger a confidencialidade, a integridade e a autenticidade de informações é o Rivest-Shamir-Adleman (RSA), que foi um dos primeiros sistemas propostos que fazem uso desse método, ainda mantendo a popularidade atualmente. O RSA baseia-se na dificuldade de fatorar grandes números primos, um problema para o qual não se conhece nenhuma solução eficiente (RIVEST, SHAMIR, ADLEMAN, 1978). Isso torna o RSA um sistema de criptografia assimétrica robusto e confiável. A criptografia de chave pública possui um papel essencial na segurança da internet. Ela é utilizada em diversas aplicações de segurança, como em autenticação e integridade de dados. Por exemplo, os certificados digitais usados em conexões HTTPS, ou seja, uma conexão HTTP feita de forma segura, se baseia em criptografia assimétrica para garantir a identidade de um servidor *web* (RESCORLA, 2001). A criptografia assimétrica também é essencial para *blockchain*, onde cada integrante da rede possui um par de chaves pública e privada e podem, por exemplo, utilizar suas chaves para assinar uma transação, a partir do uso da chave privada para criptografar a mesma, fornecendo autenticidade às transações da *blockchain*. Porém, a criptografia assimétrica possui desvantagens em relação a outros métodos de criptografia. Ela é significativamente mais lenta do que a criptografia simétrica, e, por esse motivo, em muitos sistemas práticos a criptografia assimétrica é usada para trocar uma chave simétrica, que é então usada para criptografia de dados em massa (MENEZES, VAN OORSCHOT, VANSTONE, 1996). Apesar de seus desafios, a criptografia assimétrica continua sendo crucial para a segurança cibernética moderna.

2.2 COMPUTAÇÃO DISTRIBUÍDA

A Computação Distribuída se refere a um estilo de computação no qual o processamento é distribuído em vários nós em rede. Sistemas distribuídos são sistemas computacionais na qual componentes localizados em redes se comunicam e coordenam suas ações apenas através de trocas de mensagens (COULOURIS, DOLLIMORE, KINDBERG, 2005). Ao invés da centralização em um único computador, usando o paradigma padrão da computação, as informações e funções são distribuídas por diversos dispositivos.

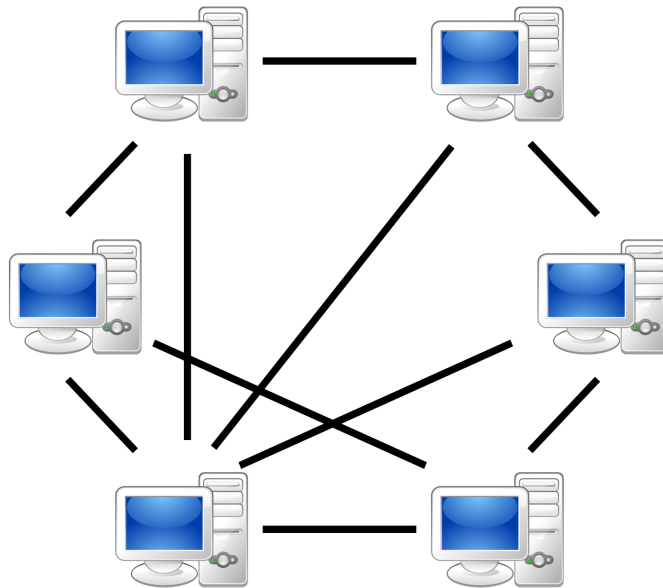


Figura 2 - Representação de uma rede computacional distribuída.

A Figura 2 representa o modelo distribuído, em que vários computadores independentes, chamados de nós, cada um com seu processamento, memória e armazenamento locais,

comunicam-se com os outros nós através de um sistema de comunicação (TANENBAUM, VAN STEEN, 2007). Essa comunicação oferece um compartilhamento de recursos vantajoso, em que se pode compartilhar tanto dados quanto poder de processamento e armazenamento. Com essa vantagem, o processamento em várias máquinas pode proporcionar redundância, possibilitando que algumas máquinas da rede parem de funcionar sem que o sistema, como um todo, pare de funcionar, proporcionando um sistema mais resiliente a falhas. Essa redundância é essencial para uma *blockchain*, onde, por exemplo, os dados das transações estão distribuídos dentre os nós participantes e ficam disponíveis para todos os membros da rede. Um aspecto-chave da computação distribuída é o *middleware*, uma camada de software que fornece um modelo de programação e oculta a heterogeneidade do *hardware* subjacente, assim não é necessário projetar o sistema para rodar em *hardware* específico, barateando o custo de implementação desses sistemas, bem como as diferenças nas linguagens de programação e no sistema operacional (EMMERICH, 2000). Isso permite aos desenvolvedores uma abstração extra, tirando deles a responsabilidade de manusear o sistema distribuído e possibilitando a concentração apenas na inovação e criação de valor, excluindo a preocupação com os detalhes de baixo nível do sistema. Apesar dessas vantagens, a computação distribuída não é isenta de desafios. Um dos principais é a coordenação entre nós, a fim de garantir o processamento correto das informações, e a garantia de consistência dos dados especialmente em situações em que várias operações acontecem simultaneamente, que podem gerar condições de corrida (LYNCH, 1996). Mesmo com esses novos problemas, com o surgimento da internet e da computação em nuvem, a computação distribuída se tornou um pilar central da computação moderna, possibilitando avanços significativos em áreas que demandam muito poder de processamento, como *big data*, aprendizado de máquina, internet das coisas e outros diversos aspectos da internet moderna.

2.3 BLOCKCHAIN

A *Blockchain*, inicialmente popularizado como a tecnologia subjacente ao *Bitcoin*, é uma tecnologia de contabilidade distribuída que permite um registro seguro, transparente e resistente à manipulação de transações (NAKAMOTO, 2008). O termo se refere ao encadeamento de blocos de um livro-razão (*ledger*) e distribuído, na qual todas as transações de uma rede *blockchain* estão localizadas, podendo ter comportamento público, como no *Bitcoin*, onde todos podem ver todas as transações, ou privado, como no *Hyperledger Fabric*, onde se tem controle de quem pode contribuir para a rede e quem acessa cada informação da rede. A segurança e a imutabilidade da *blockchain* derivam do uso de criptografia de chave pública e do conceito de mineração ou prova de trabalho e da utilização de *hashes*, garantindo integridade, autenticidade e ordenamento à *blockchain*. O termo *blockchain* traduz o próprio funcionamento da rede, em que cada bloco contém um *hash*, que é uma função matemática que mapeia dados de comprimento variável para dados de comprimento fixo, do bloco anterior, criando uma ligação contínua que torna quase impossível a alteração de blocos passados sem ser detectado (ANTONOPOULOS, 2014). Isso faz da *blockchain* uma plataforma robusta para transações seguras e transparentes. A Figura 3 é uma representação visual de como a *blockchain* do *Bitcoin* encadeia os blocos, onde o bloco mais à direita possui uma referência ao bloco anterior, proporcionando a integridade da *blockchain*. É possível visualizar o "Nonce", que é um número arbitrário que só pode ser usado uma vez, servindo como fator de aleatoriedade para formação da hash do bloco da *blockchain* do *Bitcoin*.

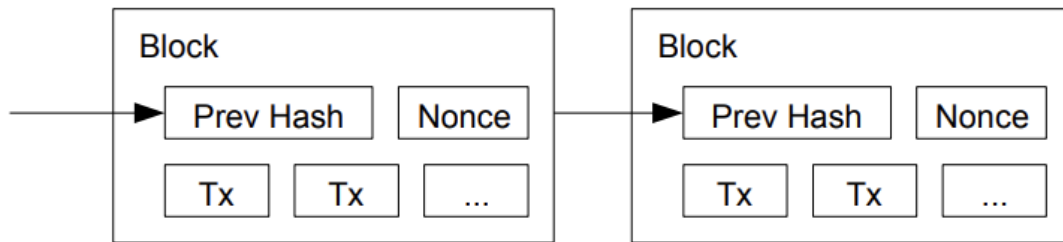


Figura 3 - Representação da blockchain.

Nesse contexto, cada bloco é uma informação de todas suas transações, armazenadas em formato de *hash*, representando a escrita (inserção à cadeia de blocos) das mesmas na rede, distribuindo essa nova confirmação para todos seus participantes. A rede é composta por uma série de computadores distribuídos, conhecidos e atuantes na rede, denominados de nós. Na rede do *Bitcoin*, esses nós são distribuídos entre nós validadores e nós mineradores. A decisão de quais blocos devem ser inseridos na rede é realizada à partir de um mecanismo de consenso, em que cada nó validador realiza uma checagem na *hash* de cada transação para validar a não ocorrência de fraude, como por exemplo o “*double spending*”, que, na rede *Bitcoin*, acontece quando um mesmo *Bitcoin* é transacionado para duas carteiras diferentes ao mesmo tempo, gerando novos *Bitcoins* no caminho e acarretando em uma fraude. Os nós mineradores são responsáveis por adicionar mais blocos à rede, através de um processo denominado prova de trabalho, na qual, por meio de um sorteio computacional, cada nó recebe um quebra-cabeças matemático complexo e o primeiro que resolver detém o direito de adicionar o novo bloco na rede. Esses nós recebem uma recompensa ao adicionar novos blocos à rede (taxa de transação). Com essa natureza descentralizada, a *blockchain* permite a verificação e registro de transações sem depender de uma autoridade central, possibilitando uma contribuição dos participantes para a obtenção de um consenso. No entanto, o *Bitcoin* também enfrenta desafios e limitações. A escalabilidade é um dos problemas principais, pois a necessidade de todos os nós validarem todas as transações pode limitar a capacidade da rede. Além disso, a natureza permanente e transparente da *blockchain* pode levantar questões de privacidade. Também existem preocupações legais e regulatórias que ainda precisam ser resolvidas (TAPSCOTT, TAPSCOTT, 2016), como o fato de não ser necessário uma autenticação prévia, onde seriam coletados dados pessoais, para usar as redes *blockchain*, como *Bitcoin* e *Ethereum*. Com isso é possível utilizar essas redes para cometer crimes financeiros, como lavagem de dinheiro e venda de produtos ilegais.

2.4 SMART CONTRACTS

Os contratos inteligentes, ou *smart contracts*, são protocolos computacionais que facilitam, verificam e aplicam a negociação e execução de um contrato, possibilitando transações confiáveis sem a necessidade de um intermediário (SZABO, 1996). Esses contratos “autônomos”, que possuem a habilidade de serem executados após uma transação de gatilho, com as regras definidas no acordo e especificadas em código, têm o potencial de aumentar a eficiência e reduzir a dependência em sistemas transacionais para cumprimento do acordo. Para o desenvolvimento de contratos inteligentes, se faz necessário a utilização de uma plataforma que suporte-os. Atualmente uma das principais plataformas é a *Ethereum*, que introduziu um ambiente computacional completo e *turing-complete*, um sistema de regras de controle de dados que pode ser usado para simular qualquer máquina de Turing, na *blockchain*, chamado de *Ethereum Virtual Machine* (EVM). Em contraste com a *Bitcoin*, que possui uma linguagem de script limitada, a *Ethereum* permite a criação e execução de contratos inteligentes complexos (BUTERIN, 2014). *Smart contracts* apresentam uma ampla

gama de potenciais aplicações, como finanças, cadeia de suprimentos e governança. Eles podem ser usados para criar aplicações descentralizadas (dApps), organizações autônomas descentralizadas (DAOs) e *tokens* não fungíveis (NFTs), entre outras aplicações (TAPSCOTT, TAPSCOTT, 2016). Porém, questões legais e regulatórias em torno dos contratos inteligentes ainda precisam ser completamente resolvidas (HARROP, KNOTTENBELT, 2018). Essas organizações autônomas descentralizadas muitas vezes operam de forma independente de nenhum Estado, o que dificulta qualquer procedimento judiciário, elevando o nível de risco ao utilizar essas soluções. Por não necessitarem de uma estrutura organizacional para administração, contratos com erros ou vulnerabilidades podem apresentar perdas irreparáveis aos detentores dos mesmo, ainda mais agravado por conta de sua imutabilidade, fazendo com que a atualização de um contrato falho não seja possível, apenas a criação de um contrato substituto, invalidando o passado, corrigindo um possível erro.

2.5 HYPERLEDGER FABRIC

Hyperledger Fabric é uma plataforma de *blockchain* privada, destinada a aplicações empresariais, desenvolvida pela IBM e *Digital Asset* em 2015 e lançada pela *Linux Foundation* em 2017. Destaca-se de outras plataformas de *blockchain* por seu desenho modular e configurável, permitindo a implementação de um alto grau de confidencialidade, resiliência, flexibilidade e escalabilidade (ANDROULAKI et al., 2018). Além disso, essa plataforma disponibiliza uma visão atual da *blockchain*, chamada de *world state*, que mitiga a necessidade de percorrer toda a cadeia de transações para encontrar a versão mais atual dos dados. O *Fabric* tem como uma das principais diferenças a outras plataformas de *blockchain* o seu suporte a contratos inteligentes escritos em linguagens de programação de propósito geral, como *JavaScript* e *Go*, e pelo uso de um mecanismo de consenso *pluggable*, ao contrário do *Ethereum* que suporta apenas linguagens mais específicas, como *Solidity*. Esta característica permite que diferentes mecanismos de consenso sejam usados dependendo das necessidades específicas da rede, sem a necessidade de um consenso de prova de trabalho intensivo em energia (CACHIN, 2016). Uma das características principais do *HyperLedger Fabric* é a sua abordagem baseada em canais para a privacidade. Comparada a outras *blockchains*, que todas as transações são visíveis para todos os participantes, o *Fabric* permite a criação de canais privados, em que apenas participantes autorizados podem ver as transações. Isso é particularmente útil para casos de uso em que a privacidade dos dados é essencial (RATHEE et al., 2020). No entanto, a configuração e manutenção de uma rede *Fabric* pode ser complexa devido à sua natureza modular, que introduz um custo maior em sua implementação. Outro desafio é a realização de atualizações ou mudanças na rede, pois isso exige uma coordenação entre todos os membros, que pode ser demorado em uma rede grande (ANDROULAKI et al., 2018).

2.6 BIG DATA

Big data refere-se ao grande volume de dados, estruturados e não estruturados, que inundam os negócios diariamente. Estes dados são caracterizados por três principais atributos, conhecidos como os 3Vs: volume, velocidade e variedade (LANEY, 2001). Com essas 3 características principais, *big data* requer um processamento diferente do tradicional, necessitando, muitas vezes, da utilização do paradigma da computação distribuída para processar esses dados de forma a extrair valor dos mesmos. Além da definição formal, também é possível classificar uma aplicação como *big data* pelo uso de ferramentas *NoSQL*, que representa soluções que não seguem o paradigma *SQL* tradicional, e pela utilização do paradigma *online analytical processing* (OLAP), ao invés do *online transaction processing* (OLTP), que é o sistema transacional mais comum, usado em plataformas *web* e em sistemas do dia-a-dia. Com isso podemos classificar nossa base de dados como *big data*, pois ela fará

uso de ferramentas *NoSQL* e utilizará o sistema OLAP. Com a crescente digitalização dos negócios e da sociedade, cada vez mais dados estão sendo criados e armazenados. Estes dados, quando adequadamente analisados, podem fornecer *insights* valiosos e orientar decisões estratégicas. Através do uso de técnicas de mineração de dados e aprendizado de máquina, padrões e conexões podem ser descobertos em conjuntos de *big data* que podem fornecer vantagem competitiva e otimização de processos (CHEN, MAO, LIU, 2014). O uso do *big data* tem aplicações em uma variedade de campos. No comércio, por exemplo, o *big data* pode ser usado para análise preditiva, otimizando as operações e permitindo uma melhor segmentação de mercado. Na saúde, o *big data* pode ser usado para melhorar os cuidados ao paciente, prever epidemias e realizar pesquisas médicas avançadas (RAGHUPATHI & RAGHUPATHI, 2014). Apesar de seu potencial, o uso do *big data* também apresenta desafios. Questões de privacidade e segurança são preocupações significativas, pois grandes volumes de dados pessoais são frequentemente coletados e analisados. Além disso, a integridade e a qualidade dos dados podem afetar a eficácia dos *insights* gerados. Assim, o gerenciamento eficaz de *big data* requer uma combinação de tecnologia, governança de dados e habilidades analíticas (KITCHIN, 2014).

2.7 METABASE

Metabase é uma ferramenta de inteligência de negócios de código aberto que facilita o acesso e a análise de informações em uma organização, servindo como plataforma para análise de *big data*. Ela permite que usuários sem conhecimento técnico em SQL realizem consultas e obtenham insights a partir de seus dados. A interface intuitiva do *Metabase* e sua capacidade de integração com várias fontes de dados tornam-no uma escolha popular para empresas que buscam democratizar o acesso a informações analíticas. Uma das principais características do *Metabase* é sua interface de usuário amigável, que permite a criação de *dashboards* e relatórios personalizados. Os usuários podem realizar consultas complexas com poucos cliques e visualizar os dados em diferentes formatos, como gráficos, tabelas e mapas. Além disso, o *Metabase* oferece recursos de filtragem e segmentação de dados, facilitando a análise específica de subconjuntos de dados. No contexto empresarial, o *Metabase* é utilizado para monitorar KPIs (*Key Performance Indicators*), realizar análises de tendências de mercado, e entender melhor o comportamento do cliente. Ele também é útil em cenários de análise de dados operacionais, como monitoramento de desempenho de vendas, análise de eficiência de processos internos e otimização de cadeias de suprimentos. Embora o *Metabase* seja uma ferramenta poderosa, ele apresenta desafios, como a necessidade de garantir a segurança dos dados e a gestão adequada de permissões de usuários. Além disso, com o avanço da tecnologia de dados, espera-se que o *Metabase* continue evoluindo, integrando mais recursos de inteligência artificial e aprendizado de máquina para fornecer análises preditivas e insights mais profundos.

3. METODOLOGIA

A presente pesquisa tem como proposta principal disponibilizar uma plataforma de consultas direcionada aos gestores políticos, como o Ministro da Educação, para que seja possível visualizar dados educacionais armazenados na *blockchain* da Jornada do Estudante. Essa plataforma idealizada se dará com uma conexão dos dados da *blockchain* ao *Metabase*. Atualmente o MEC já utiliza o *PowerBI* em alguns painéis de observabilidade, contudo o *Metabase* é uma ferramenta que apresenta uma vantagem por ser gratuito e, por causa de sua natureza de código aberto, possuir uma maior elasticidade para se conectar com os dados da *blockchain* construída sobre o *Hyperledger Fabric*. A presente pesquisa deve se integrar ao projeto Jornada do Estudante, que é mantido por uma parceria realizada entre o Laboratório de Segurança em Computação (LabSEC), o Ministério da Educação (MEC), e Laboratório

Bridge, e, de acordo com o MEC, mapeia todo o ciclo de vida da experiência educacional de um estudante, incluindo instituições educacionais frequentadas, desde a admissão até a formatura, seu desempenho e outras métricas, com uma plataforma robusta de análise de dados. O projeto Jornada do Estudante não engloba o acesso aos dados por uma ferramenta analítica, assim, existe a oportunidade da aplicação dos resultados dessa pesquisa para a integração da *blockchain* utilizada com uma plataforma de *BI* através do consumo dos dados mais recentes da *blockchain* e seu armazenamento em uma plataforma *NoSQL*. Outra questão relevante a ser analisada na implantação de uma plataforma de consulta é a gestão educacional, que é um campo de prática e estudo que se preocupa com a organização, administração, planejamento e avaliação das instituições de ensino, incluindo a liderança das pessoas e recursos para alcançar os objetivos educacionais desejados (BUSH, 2003). Com essa ferramenta que visa integrar dados educacionais por meio de uma integração de *blockchain* com *big data*, o gestor terá um maior controle na gestão educacional, que proporciona uma tomada de decisão mais concisa e com bases analíticas sólidas e extraindo um maior valor dos dados da *blockchain*.

4. DESENVOLVIMENTO

O Ministro da Educação é uma autoridade de extrema relevância ao se tratar de tomada de decisão para a área educacional, especialmente na área de planejamento acadêmico para o futuro, surgindo a necessidade de informações completas sobre a caminhada acadêmica dos alunos. Para tanto, a visão do progresso educacional pelo Ministro da Educação é custosa nos dias atuais, sendo realizada anualmente através do Censo da Educação Superior, pois é necessário coletar os dados em cada universidade a fim de centralizá-los no Inep. Cada universidade armazena seus dados de progresso educacional de forma independente, isso acarreta em um trabalho extra de normalização, padronização e condensação dos dados. Com isso, a velocidade das tomadas de decisão é executada de forma lenta. A disponibilização de um ambiente de dados normalizado, padronizado e condensado, facilita o processo de tomada de decisões pelo Ministro da Educação, visto que essa forma de dados possui uma estrutura mais limpa e acessível. A *blockchain* busca trazer esse ambiente, exigindo padrões para inserções de dados por parte do participante, garantindo a estrutura dos dados. Além disso, é possível acessar todos os dados simultaneamente, visto que todas as inserções de todos os participantes estarão disponíveis na *blockchain*, porém, no caso da *blockchain* da Jornada do Estudante, esses dados estão em conformidade com a LGPD e não podem ser acessados a nível individual, apenas através de junções ou anonimização dos dados. Com essa junção dos dados das universidades na *blockchain* da Jornada do Estudante o processo de análise de dados e tomada de decisão é facilitado devido ao novo acesso às informações, o que faz com que o Ministro da Educação tenha a sua disposição todos os dados a todo momento, favorecendo tomadas de decisões críticas e sensíveis. Esse novo formato de visualização, permite uma maior transparência nas informações, trazendo uma maneira de analisar os dados, de forma visual e já modelada, aliado com maior velocidade de atualização dos dados, os disponibilizando assim que são incluídos no *world state* da *blockchain*. A representação visual dos dados por meio de *dashboards*, permite com que sejam identificados padrões nos dados, auxiliando no processo de correção de anomalias e prevenindo erros futuros. Além de possibilitar uma monitoração constante do desempenho acadêmico dos alunos, possibilitando uma previsão de algum novo problema e, com base na centralização de informação, acaba por facilitar ações corretivas pelo Ministério da Educação e o Ministro da Educação. Fazendo o uso de tecnologias de análise de dados, é possível maximizar a ingestão de dados e minimizar o tempo para disponibilidade de um dado. Buscando a utilização de ferramentas abertas e gratuitas, é possível recorrer ao *Metabase*, devido a sua ampla comunidade, seu acesso gratuito e por ser uma plataforma de código aberto, para consultar os dados da *blockchain*. É

possível construir painéis com informações completas para serem analisadas pelos responsáveis, tornando o acesso à informação mais democrático dentro do MEC. O *Metabase* disponibiliza uma ampla configuração de permissões, que possibilita que sejam criados vários níveis de acesso, protegendo informações sensíveis de quem não deve ter acesso às mesmas. Assim, é possível realizar ajustes, de forma que o Ministro da Educação tenha o maior nível de acesso, podendo ver todas as informações disponíveis. A partir da conexão da ferramenta analítica aos dados da *blockchain* privada, através da utilização do *Metabase* e do *world state* do *Hyperledger Fabric*, é possível garantir que a informação é válida e verificada por cada membro da rede, aumentando a confiabilidade dos dados e impedindo que informações não verdadeiras, ou adicionadas de forma não convencional, impacte a avaliação do progresso dos alunos.

4.1 RESULTADOS ESPERADOS

Com a implementação da abordagem exposta na pesquisa, espera-se uma otimização no processo de tomada de decisões em relação aos cursos superiores no Brasil, buscando uma realocação de recursos mais eficientes entre as universidades, tornando mais efetivos os programas de incentivo à educação. Além disso, é esperado que exista uma nova visão em relação a formação de novos profissionais, possibilitando o adiantamento da formatura de alunos em cursos requisitados, como médicos durante uma crise sanitária ou professores de ensino básico para auxiliar em creches, permitindo uma alocação eficiente e eficaz de novos profissionais no mercado. Outro propósito dessa ferramenta é fornecer uma visão ampla do desempenho dos alunos, permitindo que ações direcionadas sejam tomadas, reduzindo a evasão dos cursos e aumentando a satisfação dos alunos para com a universidade. Em suma, observa-se que a proposta de combinar *big data* e *blockchain* para analisar dados educacionais surge da necessidade de gerenciar eficientemente dados educacionais descentralizados e padronizá-los. Esta abordagem oferece um meio mais preciso e conciso de tomar decisões baseadas em dados, fornecendo uma visão analítica abrangente dos dados educacionais. Além disso, a utilização de painéis (*dashboards*) é uma parte crucial da proposta, pois eles fornecem uma interface visual para representar e analisar os dados coletados. Os painéis facilitam o acesso a informações complexas de uma maneira mais compreensível, permitindo que os gestores educacionais tomem decisões informadas com base em análises de dados detalhadas e atualizadas. Com isso, esses painéis ajudam a identificar tendências, padrões e áreas de melhoria no sistema educacional. Essa proposta é valiosa por oferecer uma abordagem mais eficiente e baseada em dados para a tomada de decisões, melhorando a gestão educacional. A tecnologia *blockchain* fornece segurança e transparência desses dados, enquanto a centralização de informações anteriormente dispersas facilita o acesso e a análise. Além disso, ao fornecer *insights* detalhados sobre métricas educacionais, como o desempenho dos alunos, esta abordagem tem o potencial de melhorar significativamente o sistema educacional.

4.2 IMPLEMENTAÇÃO

A partir dos dados armazenados na *blockchain*, foi realizada uma extração do *world state*, que representa o estado mais atualizado da *blockchain*, recuperando os dados mais recentes de cada contrato. Com isso, esses dados foram escritos no *MongoDB*, uma ferramenta de banco de dados *open-source* gratuita que possui um formato de armazenamento similar ao *world state* do *Hyperledger Fabric*, denominado *json document*, que consiste em um documento “chave: valor”, onde cada documento representa um registro no banco de dados. Após a extração e armazenamento dos dados, foi utilizado o *Metabase* para analisar as novas tabelas e extrair informação dos dados armazenados. O *Metabase* possui diversos *drivers* para conexão com diversas bases de dados, porém não fornece suporte aos dados da *blockchain* da Jornada do Estudante. Com isso, foi necessário utilizar uma ferramenta para extração de

dados diretamente da *API (Application Programming Interface)*, que é um conjunto de rotinas e padrões que permitem que dois ou mais aplicativos se comuniquem entre si, da Jornada do Estudante e escrever esses dados no *MongoDB*.

4.2.1 Público alvo

A ferramenta é destinada para as autoridades responsáveis pela tomada de decisão dentro do Ministério da Educação, com foco no Ministro da Educação, e permite que vários servidores do Ministério tenham acesso aos painéis com permissões de visualizações em diversos níveis, protegendo dados sensíveis e especificando a função de cada um dentro do sistema. Os painéis têm como foco o Ministro da Educação pois ele é quem detém o poder de decisão dentro do MEC, decidindo quais ações serão tomadas para auxílio dos alunos. Esses painéis também possuem valor para os reitores das universidades e os diretores de centro. Essa visão detalhada do progresso dos alunos possibilita que programas de incentivo sejam executados com maior precisão e fornece uma plataforma onde é possível medir o impacto dessas medidas semestre-a-semester, aumentando ainda mais a efetividade dessas intervenções.

4.2.2 Protótipo

Para realização do protótipo, foi utilizada a rede de testes da Jornada do Estudante do MEC, disponibilizada pelo Laboratório de Segurança em Computação (LabSEC) da UFSC e preenchida com dados fictícios, com o intuito de preservar dados sensíveis. A *blockchain* de teste utilizou o *Minifabric*, que fornece um ambiente do *Hyperledger Fabric* de forma compacta e versátil, utilizando *Docker*, uma plataforma de software livre que permite aos desenvolvedores criar, implementar e executar aplicativos em contêineres, que é um pacote leve e empacotado de software que inclui tudo o que o aplicativo precisa para funcionar, incluindo código, bibliotecas e configurações, para fornecer uma solução prática para instalação local, sem necessidade de um ambiente com vastos recursos de processamento e memória. Os dados utilizados foram construídos de forma a replicar dados de produção, porém sem informações reais.

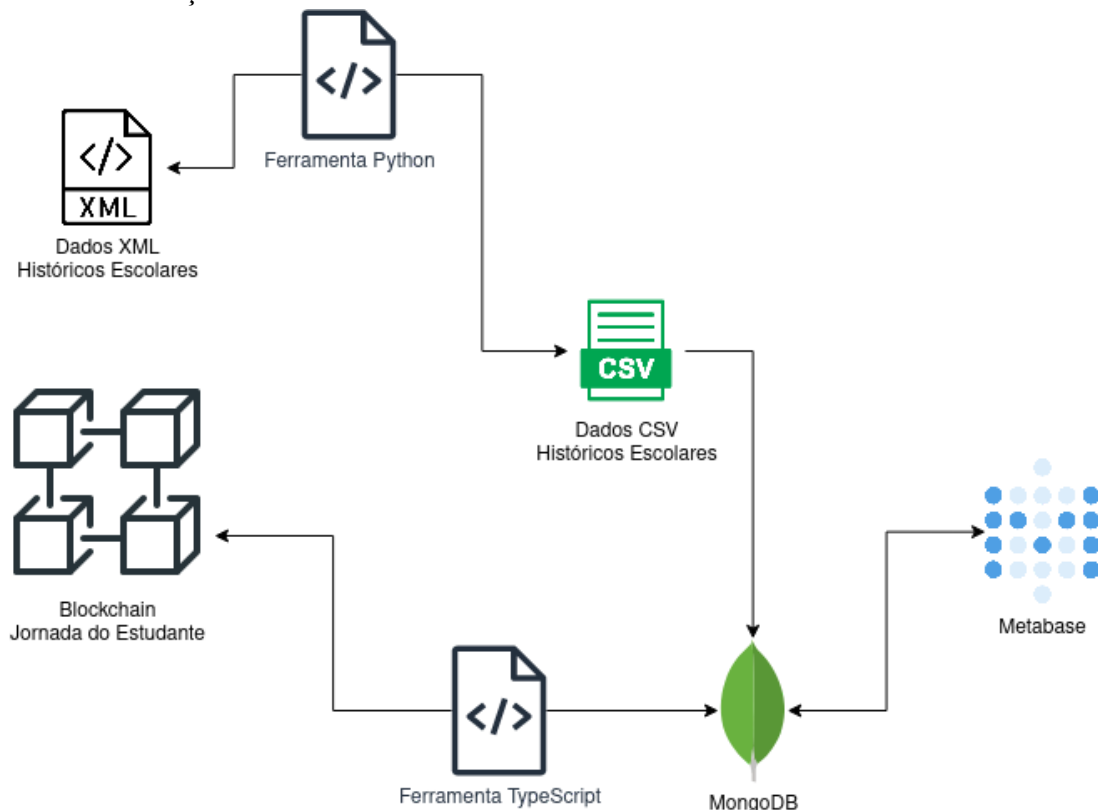


Figura 4 - Fluxograma de ingestão dos dados.

Como demonstrado na Figura 4, os dados da *blockchain* foram extraídos utilizando uma ferramenta disponibilizada pelo LabSEC, escrita em *TypeScript*, um superconjunto sintático de *JavaScript*, que fornece uma tipagem adicional à linguagem. Além disso, foi desenvolvida uma ferramenta escrita em *Python* (disponível em <https://github.com/pedronobrega/ufsc-tcc-pedro-nobrega>), para realizar a extração dos históricos escolares, armazenados em *XML (eXtensible Markup Language)*, uma linguagem de marcação que pode ser usada para representar dados de uma forma estruturada, e realizar a conversão desses dados para *CSV (Comma-separated values)*, um formato de arquivos onde os valores são separados por vírgula. Esses dados foram armazenados no *MongoDB*, pois é o banco com maior semelhança ao formato de dados do *world state* do *Hyperledger Fabric*. Utilizando o *Metabase*, foi possível acessar os dados armazenados no *MongoDB*. Com esses dados foi necessário fazer uma modelagem do banco de dados, de forma a desaninhar os dados aninhados, que são dados organizados em níveis, onde cada nível contém dados relacionados ao nível anterior, em uma única tabela que contém apenas um nível, de forma que o *Metabase* consiga acessar os dados de forma direta, sem precisar processar os dados ou construir consultas complexas para planificar os dados. Essa modelagem possibilitou a construção de painéis para visualização dos dados. Para essa pesquisa, foram escolhidos os seguintes painéis, dando foco para informações referentes aos alunos:

- Número de alunos por sexo e idade;
- Nota média por disciplina;
- Desempenho por região de origem do aluno;
- Nota média por disciplina por período.

Todos esses painéis fazem uso da tabela referente ao histórico escolar do aluno. Além destas visualizações, é possível filtrar essas informações nos seguintes níveis:

- UF (Unidade Federativa);
- IES (Instituição de Ensino Superior);
- Curso;
- Disciplinas;
- Período letivo;
- Sexo.

Com essa granularidade, é possível investigar como está o progresso dos alunos, observando, por exemplo, uma alta taxa de reprovação em certa disciplina, ao comparar o desempenho médio entre os períodos. Outra possibilidade é realizar uma análise demográfica dos alunos, comparando a nota média nas disciplinas em relação à UF de origem do aluno, possibilitando, por exemplo, uma visualização do impacto de estudar em uma universidade tão distante de sua família. Com as visões de dados atualizados em formato de gráficos é possível analisar de forma profunda como está o desempenho dos alunos nos cursos, sendo transparente em vários fatores, como o desempenho de alunas do sexo feminino em cursos que apresentam uma maioria masculina e vice-versa, assim como é possível visualizar o impacto de reprovações em disciplinas importantes no curso, como disciplinas que são pré-requisito de várias outras no decorrer do currículo, e entender como mitigar possíveis consequências dessas reprovações. Com base nessas informações, é possível realizar filtros específicos, observando estas estatísticas em um nível ainda mais granular, selecionando os dados por sexo do aluno ou pelo seu período de ingresso, auxiliando a visão completa de como certos grupos estão desempenhando sua graduação.

5. RESULTADOS E DISCUSSÃO

Com base na implementação dos painéis e na utilização de dados de exemplo, seguindo o formato dos dados reais da *blockchain*, foram obtidos os diversos gráficos para demonstrar os dados de uma forma clara e concisa.

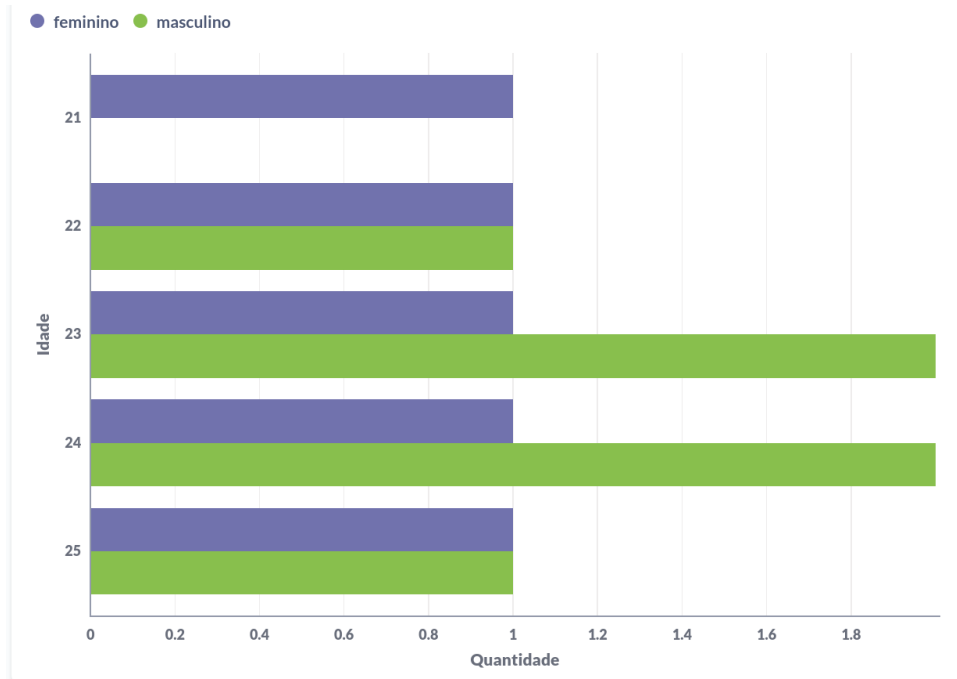


Figura 5 - Alunos por sexo e idade.

No gráfico apresentado na Figura 5 é possível ter uma visão da distribuição demográfica dos alunos, para visualizar a quantidade de alunos por idade e sexo. Esses dados podem ser filtrados por alguns atributos desses alunos, como o curso, a disciplina e o período letivo que a disciplina foi cursada. Com base no gráfico da Figura 5, é possível verificar como está a distribuição dentro de cada universidade, fazendo com que seja possível direcionar esforços para fornecer oportunidades iguais para alunos de qualquer idade e qualquer sexo, fornecendo oportunidades iguais para todos.

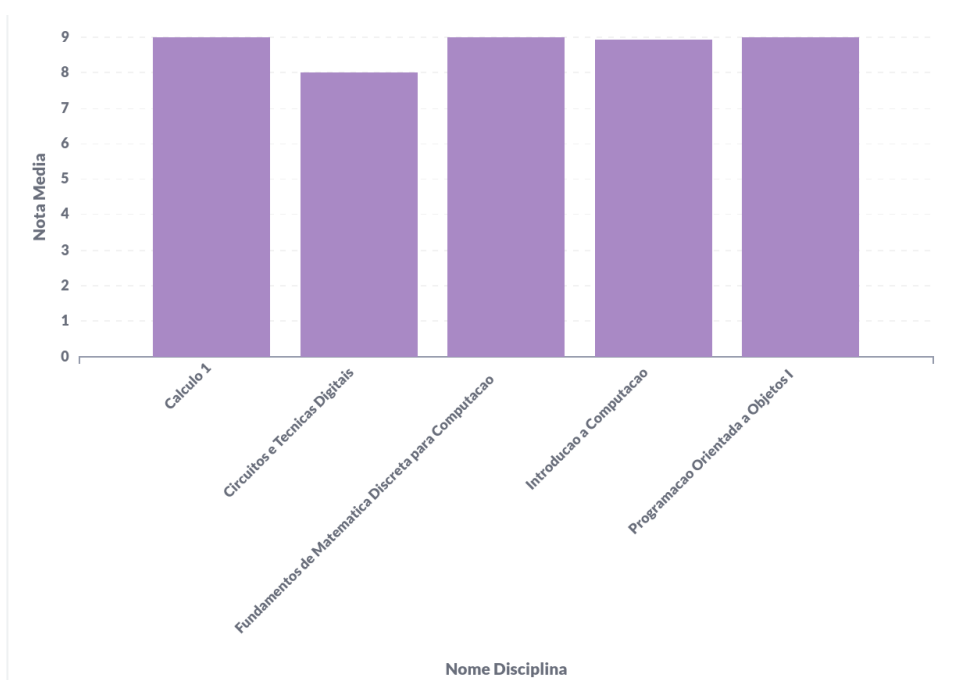


Figura 6 - Nota média por disciplina.

Com base na visualização do gráfico apresentado na Figura 6, podemos observar a média das notas dos alunos de cada disciplina, verificando o desempenho dos mesmos nas disciplinas cursadas. Como demonstra a Figura 6, o gráfico permite o filtro pelo curso, período letivo e o sexo dos alunos, assim é possível compreender a performance dos estudantes conforme avançam no curso. Por meio de gráficos com uma visibilidade granular, é possível identificar pontos de inflexão no curso, ocasionando em desistências. Assim, pode-se direcionar esforços para corrigir como as disciplinas são ministradas e/ou avaliadas, aumentando a taxa de permanência no curso.

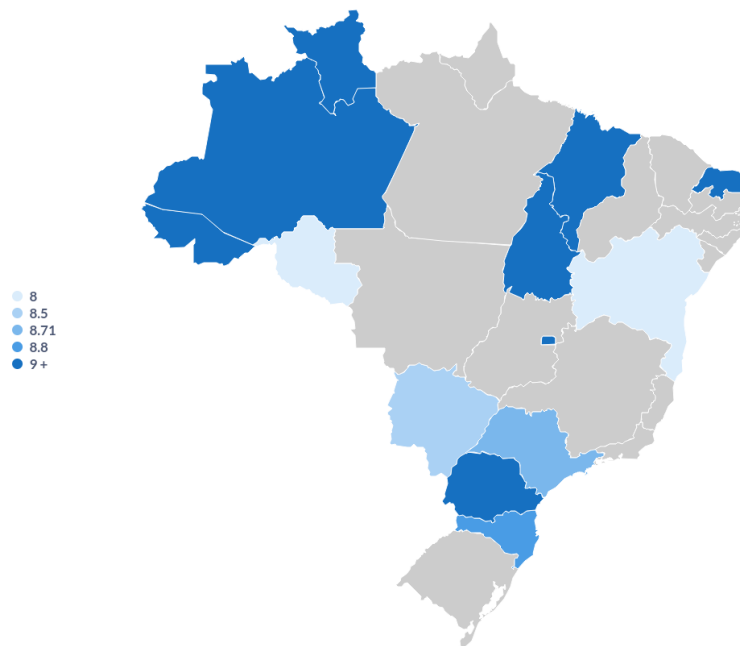


Figura 7 - Desempenho por estado de origem.

O gráfico exibido na Figura 7 apresenta uma visualização utilizando mapa de calor e demonstrando a nota média pelo estado de origem de cada estudante, podendo ser filtrado por curso, disciplina, período letivo e sexo do aluno, esse gráfico demonstra o desempenho acadêmico dos alunos de forma granular, podendo observar como cada grupo se desempenha no decorrer do percurso acadêmico. A visualização utilizando um mapa de calor é útil para compreender como auxiliar alunos que mudam sua residência com o intuito de cursar uma universidade em outro estado e acabam por romper sua rede de apoio familiar, o que pode ocasionar em uma diminuição das notas e um acréscimo na taxa de evasão do curso.

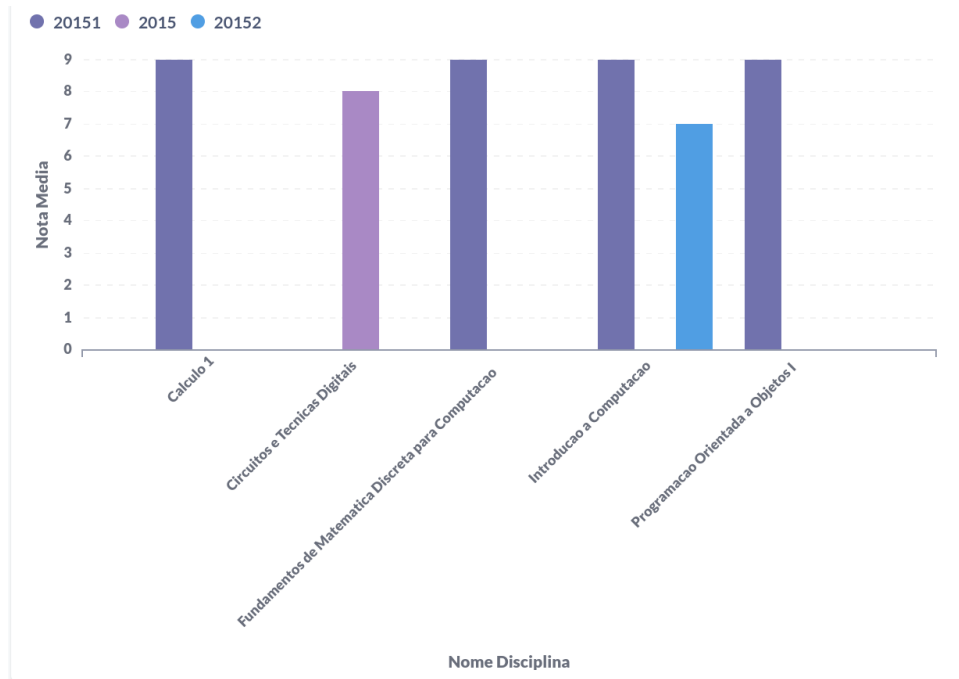


Figura 8 - Nota média das disciplinas por período.

A visualização apresentada no gráfico da Figura 8 fornece um entendimento de como cada disciplina está sendo aproveitada pelos alunos ao longo do tempo. Podendo ser filtrada por sexo do aluno e por períodos letivos específicos, é possível acompanhar a efetividade das metodologias aplicadas em cada disciplina. Com o gráfico da Figura 8 é possível observar possíveis dificuldades na passagem de conhecimento das disciplinas, acompanhar a mudança de metodologias e/ou professores responsáveis. Assim, é possível planejar como as disciplinas são ministradas, com o intuito de aumentar o aproveitamento estudantil e a taxa de permanência nos cursos. Com base nos resultados obtidos, é possível concluir que esses gráficos apresentam uma visão completa do que pode ser feito com os dados da *blockchain* do Jornada do Estudante, fornecendo informações valiosas de como os alunos estão aproveitando a universidade e os cursos ofertados. Podemos analisar também o papel do *Metabase* na construção dos gráficos. Essa ferramenta possui um controle de acesso à cada base de dados, podendo ser configurada quem tem acesso a cada um dos gráficos e das *dashboards*, assim preservando a privacidade dos estudantes e disponibilizando para os funcionários do MEC apenas os gráficos relevantes para sua função. Porém, alguns pontos devem ser levantados, como a falta de possibilidade de extração dos gráficos, que permitiria fazer backup externo dos mesmos, assim como a falta de uma forma direta de consulta à *json documents*, o formato adotado pelo *world state* do *Hyperledger Fabric* e do *MongoDB*, na qual não é possível consultar os dados aninhados, sendo necessário um trabalho de planificação dos dados e a criação de novas tabelas para consulta dos mesmos. Outro ponto importante é a falta de possibilidade de criação de gráficos complexos com maior dinamismo com relação aos filtros, que possibilitam uma visualização mais completa e com maior valor agregado.

5.1 CONSIDERAÇÕES FINAIS

Ao longo desta pesquisa, foi possível observar a importância da utilização de tecnologias como *big data* e *blockchain* na análise de dados educacionais. Por meio da análise de dados de teste, já padronizados, do Jornada do Estudante, foi possível identificar pontos de inflexão no curso, ocasionando em desistências, e direcionar esforços para corrigir como as disciplinas são ministradas e/ou avaliadas, aumentando a taxa de permanência no curso. Além disso, a utilização da tecnologia *blockchain*, aliada ao controle de acessos do *Metabase*, permitiu a

segurança e transparência dos dados educacionais, garantindo que apenas pessoas autorizadas tenham acesso a essas informações. Isso é especialmente importante em um contexto em que a privacidade e a segurança dos dados são cada vez mais valorizadas. Por fim, é importante destacar que os resultados obtidos nesta pesquisa podem ser utilizados para melhorar a qualidade da educação no Brasil. Ao compreender a performance dos estudantes conforme avançam no curso, é possível identificar as disciplinas que apresentam maiores dificuldades e direcionar esforços para melhorar o ensino dessas matérias. Dessa forma, espera-se contribuir para a formação de profissionais mais capacitados e preparados para enfrentar os desafios do mercado de trabalho.

6. CONCLUSÃO

Em suma, a presente pesquisa apresentou um sistema inovador para análise de dados educacionais utilizando tecnologias como o *Metabase* e a *blockchain* do projeto Jornada do Estudante. Através da utilização dessas tecnologias, foi possível garantir a segurança e consistência dos dados, bem como identificar pontos de inflexão nos cursos que podem ser corrigidos para melhorar a qualidade da educação. Ao conectar o *Metabase* à base de dados da *blockchain*, foi possível modelar os dados a fim de construir novas tabelas que satisfizeram a necessidade de consulta de dados, fornecendo painéis de visualizações complexos e com uma variedade de informação. Com base nesses painéis, é possível ter uma visão do que se pode alcançar com esses dados e como os profissionais da educação podem utilizar essas informações no processo de tomada de decisão dentro do Ministério da Educação. Os resultados obtidos neste trabalho são promissores e podem ser utilizados para melhorar a qualidade da educação no Brasil. Ao compreender a performance dos estudantes e identificar as disciplinas que apresentam maiores dificuldades, é possível direcionar esforços para melhorar o ensino dessas matérias e, assim, formar profissionais mais capacitados e preparados para enfrentar os desafios do mercado de trabalho.

6.1 TRABALHOS FUTUROS

Diante dos resultados apresentados na pesquisa, surge a necessidade de continuação do desenvolvimento com projetos na mesma área. Uma oportunidade de trabalho futuro a nível de mestrado é a conexão direta da *blockchain* com o *Metabase*, criando um novo conector a fim de lidar com todas as chamadas necessárias para visualizar os dados de forma ágil e completa. Outro possível trabalho futuro que trará um valor agregado para a solução proposta, é um projeto que busca adicionar observabilidade à *blockchain* e à plataforma analítica, extraindo informações do uso no dia-a-dia, assim como garantindo a integridade e segurança dos dados. Por fim, é importante destacar que este trabalho é apenas o começo de uma jornada que pode levar a avanços significativos na área de educação. A utilização de tecnologias como *big data* e *blockchain* pode revolucionar a forma como os dados educacionais são coletados e analisados, permitindo uma compreensão mais profunda do processo de ensino e aprendizagem. Esperamos que este trabalho possa inspirar outros pesquisadores a explorar ainda mais essas tecnologias e contribuir para a melhoria da educação no Brasil e no mundo.

REFERÊNCIAS

ANDROULAKI, E. et al. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. In: Proceedings of the Thirteenth EuroSys Conference (EuroSys '18). New York: Association for Computing Machinery, Article 30, p. 1–15, 2018.

ANTONOPOULOS, A. M. Mastering Bitcoin: Unlocking Digital Cryptocurrencies. Sebastopol: O'Reilly Media, 2014.

- BUSH, T. Theories of Educational Leadership and Management. London: Sage, 2003.
- BUTERIN, V. Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform. Ethereum White Paper, 2014.
- CACHIN, C. Architecture of the Hyperledger Blockchain Fabric. In: Workshop on Distributed Cryptocurrencies and Consensus Ledgers, 2016.
- CHEN, M.; MAO, S.; LIU, Y. Big Data: A Survey. Mobile Networks and Applications, v. 19, n. 2, p. 171–209, 2014.
- COULOURIS, G.; DOLLIMORE, J.; KINDBERG, T. Distributed Systems: Concepts and Design. London: Pearson Education, 2005.
- EMMERICH, W. Software engineering and middleware: a roadmap. In: Proceedings of the Conference on the Future of Software Engineering, Limerick, Ireland, 2000.
- HARROP, W.; KNOTTENBELT, W. The DAO Attack: Mathematical Modelling and Lessons Learned. In: Gervais A., Livshits B. (eds) Financial Cryptography and Data Security. FC 2018. Lecture Notes in Computer Science, vol 10958. Cham: Springer, 2018.
- KITCHIN, R. Big Data, new epistemologies and paradigm shifts. Big Data & Society, v. 1, n. 1, 2014.
- LANEY, D. 3D data management: Controlling data volume, velocity, and variety. META Group Research Note, n. 6, 2001.
- LÜCK, Heloísa. A evolução da gestão educacional a partir de mudança paradigmática. Gestão em Rede, [s.l.], n. 3, p. 13-18, nov. 1997.
- LYNCH, N. Distributed Algorithms. San Francisco: Morgan Kaufmann Publishers Inc., 1996.
- MENEZES, A.; VAN OORSCHOT, P.; VANSTONE, S. Handbook of Applied Cryptography. Boca Raton: CRC Press, 1996.
- NAKAMOTO, S. Bitcoin: A Peer-to-Peer Electronic Cash System, 2008.
- NETO, N. N. Criptografia em Comunicação de Dados. São Paulo, SP: 2002.
- RAGHUPATHI, W.; RAGHUPATHI, V. Big data analytics in healthcare: promise and potential. Health Information Science and Systems, v. 2, n. 1, p. 3, 2014.
- RATHEE, G.; SHARMA, H.; SHARMA, V.; IYER, S. Performance Benchmarking and Optimizing Hyperledger Fabric Blockchain Platform. In: 2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID), 2020.
- RESCORLA, E. SSL and TLS: Designing and Building Secure Systems. Boston: Addison-Wesley Professional, 2001.

RIVEST, R.; SHAMIR, A.; ADLEMAN, L. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM*, v. 21, n. 2, p. 120-126, 1978.

SCHNEIER, B. *Applied Cryptography: Protocols, Algorithms, and Source Code in C*. New York: Wiley, 1996.

SAGIROGLU, S.; SINANC, D. Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, p. 42-47, 2013.

SZABO, N. *Smart Contracts: Building Blocks for Digital Markets*. *Entropy*, n. 16, 1996.

TANENBAUM, A. S.; VAN STEEN, M. *Distributed Systems: Principles and Paradigms*. Upper Saddle River: Pearson Prentice Hall, 2007.

TAPSCOTT, D.; TAPSCOTT, A. *Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business, and the World*. New York: Portfolio, 2016.

YAGA, Dylan et al. *Blockchain technology overview*. arXiv preprint arXiv:1906.11078, 2019.