

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

Trabalho Final (TF)

Big Data (BD)

Yan Andrade de Sena - 180145363

Brasília, DF

2023

a) Definição da Tecnologia Pesquisada;

Big Data, que pode ser traduzido para o português como "Grande Dados", refere-se à imensa quantidade de informações que são coletadas diariamente de diversas fontes, tais como websites, redes sociais e outras. É importante ressaltar que Big Data não se limita apenas à extração e ao armazenamento desses dados, mas também engloba sua análise e gerenciamento. Esse conjunto de informações volumoso e complexo apresenta desafios adicionais em relação à sua organização, processamento e interpretação. A análise de Big Data envolve a utilização de ferramentas e técnicas avançadas para extrair insights valiosos, identificar padrões, prever tendências e tomar decisões estratégicas com base nessas informações. Trata-se de um campo interdisciplinar que combina conhecimentos de ciência de dados, estatística, inteligência artificial e outras áreas para extrair o máximo valor desses dados abundantes e diversificados.

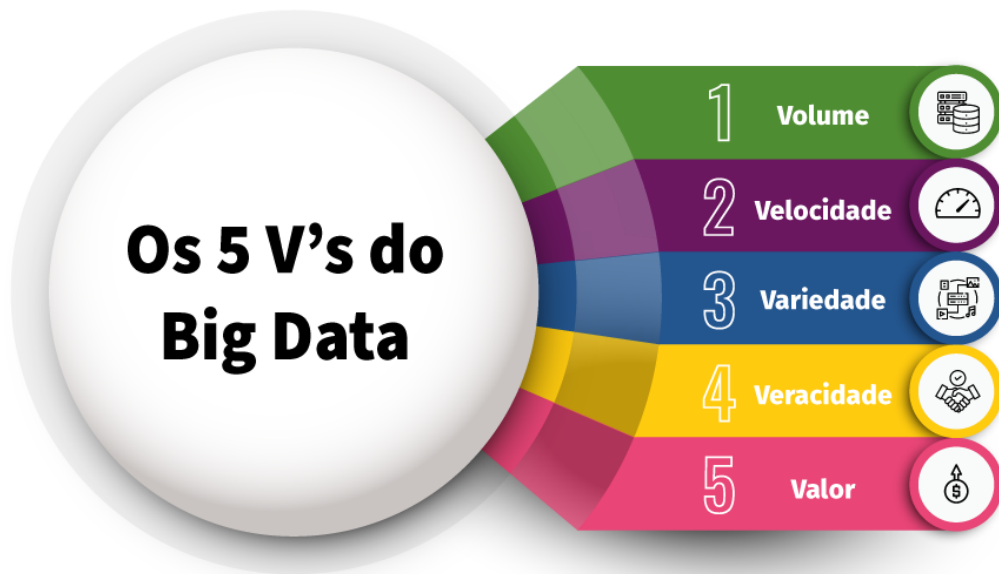
A caracterização do Big Data pode ser feita considerando inicialmente três aspectos dos dados: volume, variedade e velocidade, conhecidos como os "três Vs" da tecnologia. O volume refere-se à quantidade de dados envolvidos. É interessante destacar que, de acordo com a IDC, a quantidade de informações armazenadas dobra a cada dois anos, estimando-se que até 2025 haverá cerca de 180 zettabytes de armazenamento, o que equivale a 180 bilhões de terabytes.

Quanto à variedade de dados, ela vai além das informações transacionais e engloba dados obtidos de diversas fontes, como páginas da web, índices de pesquisa, fóruns, mídias sociais, e-mails, IoT, áudio, vídeo, entre outros. Portanto, é necessário utilizar tecnologias, algoritmos e técnicas especializadas na análise tanto de dados estruturados quanto não estruturados.

A velocidade refere-se à rapidez com que os dados são gerados, recuperados, armazenados e acessados, podendo ser entendida também como o fluxo de dados. Além dessas características destacadas, existem ainda mais dois "Vs" adicionais: veracidade e valor dos dados obtidos.

A veracidade diz respeito à confiabilidade dos dados, exigindo qualidade, consistência e uma fonte conhecida, além de não serem fabricados. Por fim, o valor é subjetivo, pois depende de todas as outras características mencionadas, sendo demandado por um objetivo específico de análise e sua aplicabilidade. Em resumo, o valor dos dados está relacionado à capacidade de gerar benefícios, dependendo

da equipe de cientistas, analistas e engenheiros de dados envolvida no processo.



Na Big Data, a organização e armazenamento de diferentes tipos de dados são realizados por meio de estruturas flexíveis e adaptáveis. A natureza diversa dos dados exige abordagens especiais para lidar com eles de forma eficiente. Aqui estão algumas das técnicas e tecnologias comumente usadas:

- **Data Lakes:** Os data lakes são repositórios que permitem armazenar uma variedade de tipos de dados, incluindo dados estruturados, semi-estruturados e não estruturados. Eles permitem que os dados sejam armazenados em seu formato original, sem a necessidade de transformação prévia. Essa abordagem flexível permite uma análise mais abrangente e exploratória dos dados.
- **Schemas flexíveis:** Ao contrário dos bancos de dados tradicionais, que exigem esquemas rígidos e estruturas fixas, na Big Data, os dados podem ser armazenados em esquemas flexíveis. Isso significa que diferentes tipos de dados podem coexistir na mesma estrutura, permitindo acomodar dados heterogêneos.
- **NoSQL:** Os bancos de dados NoSQL (Not Only SQL) são amplamente utilizados na Big Data para lidar com diferentes tipos de dados. Esses bancos de dados são projetados para serem escaláveis e flexíveis, permitindo o armazenamento e recuperação eficiente de dados não estruturados ou semiestruturados, como documentos, gráficos e dados em formato de chave-valor.
- **Processamento distribuído:** Para lidar com grandes volumes de dados, a Big Data faz uso de tecnologias de processamento distribuído, como o Hadoop e o Spark. Essas tecnologias distribuem o processamento em clusters de computadores, permitindo que dados sejam processados em paralelo, aumentando a eficiência e o desempenho.

- Indexação e metadados: Para facilitar a recuperação e a busca eficiente de dados, são utilizadas técnicas de indexação e metadados. Os dados podem ser indexados por diferentes atributos e marcados com metadados descritivos, o que facilita a localização e a análise posterior.

Essas são apenas algumas das abordagens e tecnologias utilizadas para organizar e armazenar diferentes tipos de dados na Big Data. A escolha da estratégia depende da natureza dos dados, dos requisitos de análise e das necessidades específicas do projeto ou da organização.

b) Objetivo(s) principal(is) da Tecnologia Pesquisada;

O objetivo da Big Data é aproveitar o potencial dos grandes volumes de dados disponíveis para obter insights valiosos e tomar decisões estratégicas. Ao analisar e interpretar os dados de forma inteligente, as organizações podem identificar padrões, tendências e correlações que fornecem informações úteis para aprimorar suas operações e impulsionar o crescimento. Além disso, a análise de Big Data pode auxiliar na detecção de problemas, identificação de oportunidades de mercado, personalização de experiências para os clientes e otimização de processos.

Outro objetivo fundamental da Big Data é permitir uma tomada de decisão baseada em dados sólidos e confiáveis. Ao coletar, armazenar e analisar informações de várias fontes, as empresas podem obter uma compreensão mais completa e precisa do cenário em que operam. Essa abordagem baseada em dados fornece uma base sólida para a tomada de decisões informadas, reduzindo a dependência de intuições subjetivas e aumentando a probabilidade de sucesso em várias áreas, como marketing, operações, finanças e gestão de recursos humanos.

Por fim, a Big Data busca explorar o potencial dos dados para inovação e criação de valor. Ao analisar e compreender os padrões emergentes, as empresas podem identificar oportunidades de negócios inovadoras, desenvolver novos produtos e serviços, melhorar a eficiência operacional e antecipar as necessidades e preferências dos clientes. A utilização estratégica dos dados permite que as organizações se adaptem rapidamente às mudanças do mercado e se destaquem da concorrência, impulsionando o crescimento e a sustentabilidade a longo prazo.

c) Vantagens da Tecnologia Pesquisada;

O Big Data oferece uma série de vantagens para as organizações. As principais vantagens são:

- Tomada de decisões informadas: O Big Data fornece informações valiosas que ajudam na tomada de decisões estratégicas. Ao analisar grandes volumes de dados provenientes de várias fontes, as organizações podem obter insights precisos e baseados em evidências, permitindo que tomem decisões informadas e fundamentadas.

- **Identificação de padrões e tendências:** A análise de Big Data permite identificar padrões, tendências e correlações em um nível que seria difícil ou impossível de se observar por meio de métodos tradicionais. Isso possibilita uma compreensão mais profunda do comportamento do mercado, preferências do cliente, mudanças nas demandas e outras informações relevantes que podem impulsionar a inovação e o crescimento dos negócios.
- **Melhoria da eficiência operacional:** Ao utilizar o Big Data, as empresas podem otimizar seus processos e operações. A análise dos dados pode revelar gargalos, ineficiências e oportunidades de melhoria em diferentes áreas, permitindo que sejam feitos ajustes e mudanças que resultem em maior eficiência, redução de custos e aumento da produtividade.
- **Personalização e experiência do cliente:** Com o Big Data, as empresas podem entender melhor as preferências e necessidades dos clientes. Ao analisar dados comportamentais, histórico de compras, interações nas redes sociais e outras informações relevantes, é possível oferecer experiências personalizadas, recomendações precisas e atender às expectativas dos clientes de maneira mais eficaz, o que pode resultar em maior satisfação e fidelidade.
- **Deteção de fraudes e segurança:** O Big Data pode ser usado para identificar atividades fraudulentas e detectar anomalias em tempo real. Ao analisar grandes volumes de dados, é possível identificar padrões suspeitos e comportamentos incomuns, ajudando a prevenir e mitigar riscos de segurança e fraude.

Essas são apenas algumas das vantagens do Big Data. No entanto, é importante ressaltar que a implementação eficaz do Big Data requer a adoção de tecnologias adequadas, a garantia da qualidade dos dados e a capacidade de interpretar e extrair insights relevantes dessas informações vastas e complexas.

d) Desvantagens da Tecnologia Pesquisada;

Embora o Big Data ofereça várias vantagens, também existem algumas desvantagens a serem consideradas. As principais desvantagens do Big Data são:

- **Custo de implementação:** A infraestrutura necessária para lidar com grandes volumes de dados pode ser cara. O investimento em hardware, software e recursos humanos qualificados pode ser significativo, tornando a implementação e manutenção do Big Data um desafio financeiro para algumas organizações.
- **Privacidade e segurança dos dados:** Com a coleta e análise de grandes quantidades de dados, surgem preocupações com a privacidade e a segurança. A proteção dos dados pessoais e sensíveis torna-se uma preocupação crítica, uma vez que a exposição ou o vazamento dessas informações pode resultar em danos reputacionais, problemas legais e perda de confiança dos clientes.

- Complexidade do gerenciamento e análise: O Big Data é caracterizado por sua complexidade e diversidade de fontes de dados. Lidar com a variedade de formatos, estruturas e qualidade dos dados pode ser um desafio significativo. Além disso, a análise de grandes volumes de dados requer habilidades especializadas em ciência de dados e análise estatística, o que pode ser escasso no mercado de trabalho.
- Viés e interpretação incorreta: A análise de Big Data pode levar a resultados tendenciosos ou interpretações incorretas se os dados forem coletados ou tratados de maneira inadequada. É fundamental garantir a qualidade dos dados e aplicar técnicas adequadas de análise para evitar conclusões errôneas ou enviesadas, que podem levar a decisões equivocadas.
- Legalidade e conformidade regulatória: A coleta e o uso de dados em grande escala estão sujeitos a leis e regulamentos específicos, como a Lei Geral de Proteção de Dados (LGPD) no Brasil e o Regulamento Geral de Proteção de Dados (GDPR) na União Europeia. As organizações precisam estar em conformidade com essas regulamentações e garantir que suas práticas de coleta e tratamento de dados estejam em conformidade, o que pode exigir recursos adicionais e processos específicos.

É importante considerar essas desvantagens e mitigar seus impactos por meio de políticas e práticas adequadas de governança de dados, segurança cibernética e conformidade legal. O Big Data deve ser implementado com responsabilidade e consideração aos riscos associados, garantindo que os benefícios superem as desvantagens.

e) Exemplo(s) de uso interessantes em empresas, projetos ou instituições dessa tecnologia de Banco de Dados pesquisada;

O Grupo Pão de Açúcar possui um sistema de fidelização de clientes chamado Clube Extra, que utiliza a tecnologia Big Data para promover o relacionamento não apenas com os clientes, mas também com os fornecedores. Ao se cadastrar no programa, os clientes têm a oportunidade de acumular pontos por meio de compras realizadas tanto em lojas físicas quanto online. Os dados gerados por essa plataforma são analisados para estabelecer conexões entre os clientes e os produtos, marcas favoritas e itens mais consumidos.

Graças ao uso do Big Data, o sistema operacional otimiza o controle de estoque. Isso ocorre porque a empresa tem acesso a informações prévias sobre a demanda esperada de determinado produto, o que permite tomar decisões mais precisas sobre o nível de estoque necessário, considerando também os custos envolvidos na sua manutenção.

Um exemplo de caso de insucesso envolvendo o uso do Big Data ocorreu com a empresa de varejo norte-americana Target em 2012. A empresa utilizava análises de dados para segmentar seus clientes e enviar ofertas personalizadas

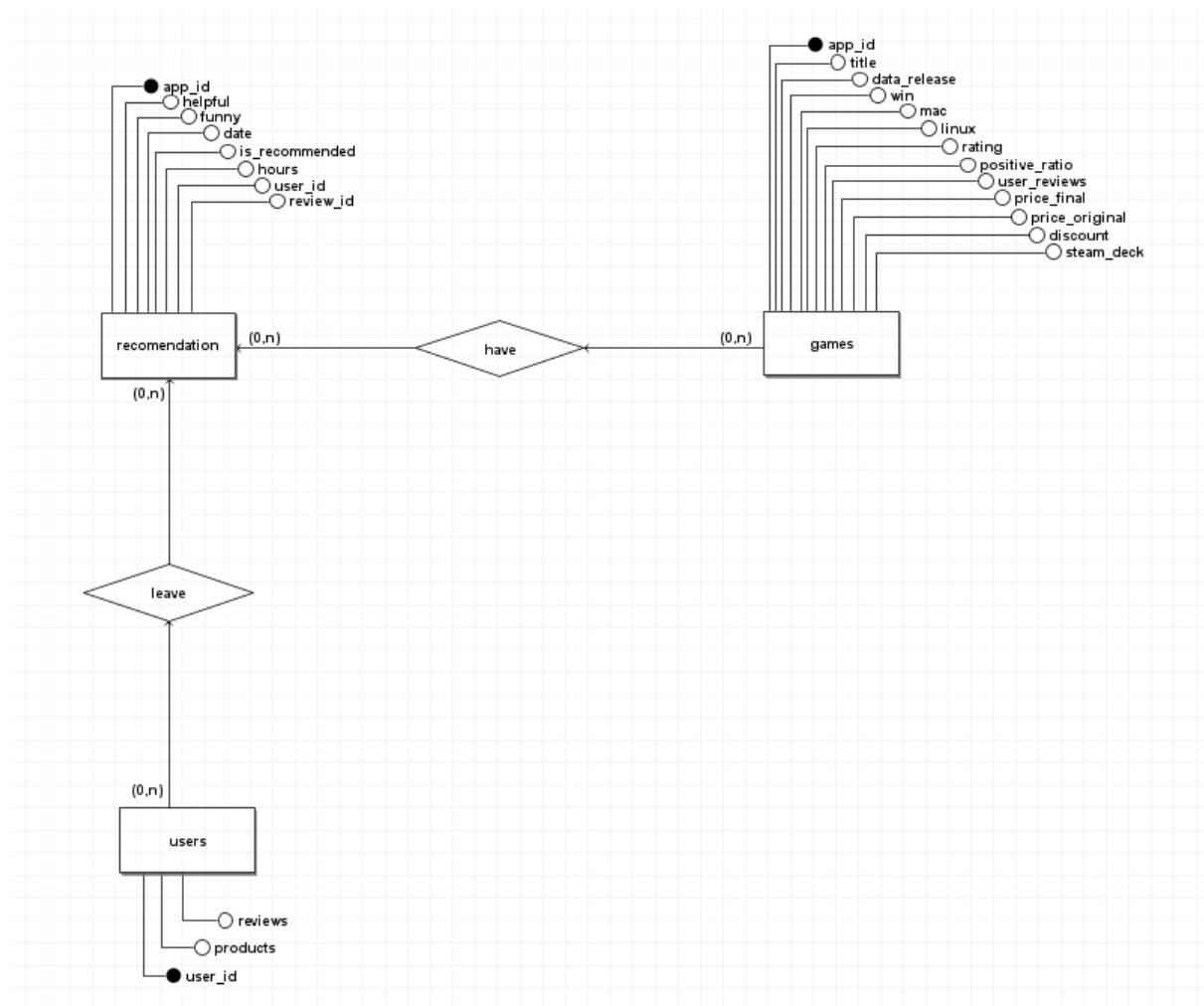
com base em seus padrões de compra. No entanto, um incidente revelou que essa estratégia nem sempre é bem-sucedida.

A Target começou a enviar cupons de desconto relacionados a produtos para mulheres grávidas com base em algoritmos que identificam padrões de compra comuns entre as gestantes. O problema ocorreu quando uma adolescente, cujos hábitos de compra sugeriam que ela estava grávida, recebeu esses cupons em sua casa. Isso acabou revelando sua gravidez para a família, causando constrangimento e revolta.

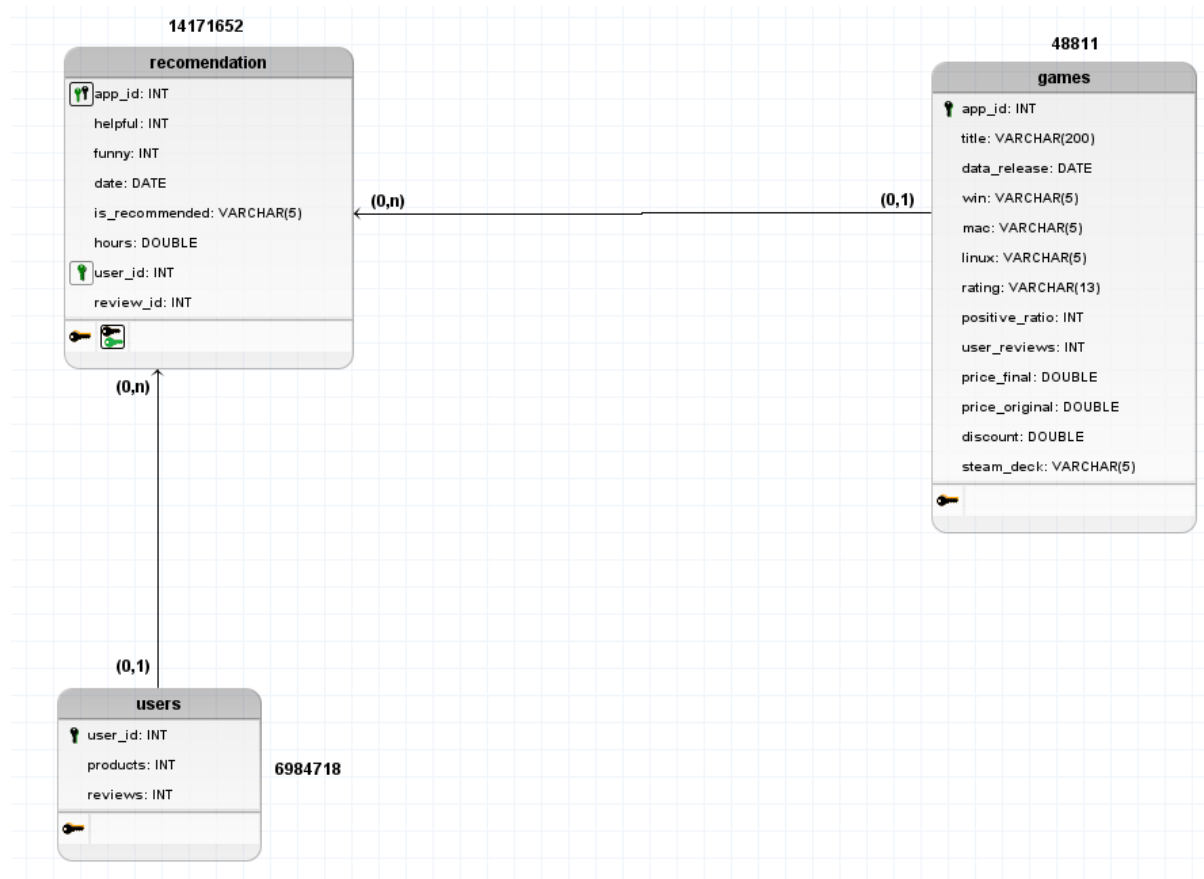
O caso da Target demonstra como o uso inadequado de análises de Big Data pode levar a resultados indesejados. A empresa presumiu que os algoritmos eram precisos e falhou em considerar a importância da sensibilidade dos dados e da privacidade dos indivíduos. O incidente ressalta a necessidade de equilibrar o uso do Big Data com a ética, privacidade e compreensão completa do contexto em que os dados são aplicados.

g) Base de Dados (documentação)

DE-R



DLD



A base de dados utilizada é a base de recomendações de jogos na plataforma Steam. Que pode ser acessada através deste link:

<https://www.kaggle.com/datasets/antonkozyriev/game-recommendations-on-steam?select=games.csv>

f) Bibliografias Pesquisadas

Referências:

Oracle. **O que é Big Data?**. Acesso em 12 de junho de 2023, disponível em: <https://www.oracle.com/br/big-data/what-is-big-data/>

Natanael Galdino. **Big Data: Ferramentas e Aplicabilidade**. AEDB. <https://www.aedb.br/seget/arquivos/artigos16/472427.pdf>

Universidade de Caxias do Sul. **Big Data: O que é, para que serve, como aplicar e exemplos**. Acesso em 12 de junho de 2023, disponível em: <https://ead.ucs.br/blog/big-data#/>

Under. **Conheça 6 casos de sucesso de Big Data em empresas**. Acesso em 12 de junho de 2023, disponível em: <https://blog.under.com.br/6-casos-de-sucesso-do-uso-do-big-data-por-empresas>

Guide Investimentos. **Big Data: Como a Target descobriu uma gravidez antes da família?**. Acesso em 12 de junho de 2023, disponível em: <https://conteudos.guide.com.br/textos/big-data-como-a-target-descobriu-uma-gravidez-antes-da-propria-familia/>

Accurate Software. **Big Data e Inteligência Artificial: o poder das análises.** Acesso em 12 de junho de 2023, disponível em: <https://blog.accurate.com.br/big-data-inteligencia-artificial/>

TAURION, Cezar. **Big Data**. Rio de Janeiro: Brasport, 2013.

Dataside. **Sistemas e formas de armazenamento de dados**. Acesso em 12 de junho de 2023, disponível em: <https://www.dataside.com.br/post/sistemas-e-formas-de-armazenamento-de-dados>