

UNIVERSIDADE DE BRASÍLIA
Faculdade do Gama

Sistemas de Banco de Dados 2

Trabalho Final (TF)
Bancos de Dados Semiestruturados

Antonio Rangel Chaves - 180098021
Artur Vinícius Dias Nunes - 190142421

Brasília, DF
2023

a) Definição da Tecnologia Pesquisada

Segundo Abiteboul, Buneman e Suciu (2014), os bancos de dados semi estruturados são sistemas de gerenciamento de informações que armazenam dados organizados de forma parcialmente estruturada. Diferentemente dos bancos de dados relacionais tradicionais, nos quais os dados são armazenados em tabelas com esquemas rígidos e relacionamentos pré-definidos, os bancos de dados semi estruturados permitem a flexibilidade na estrutura dos dados.

Os bancos de dados semi estruturados são caracterizados pela presença de dados não tabulares, como documentos XML, JSON ou mesmo texto livre. Esses dados podem ser organizados em formatos hierárquicos, como árvores, ou em formatos de grafo. Isso permite que as informações sejam armazenadas de maneira mais flexível, facilitando a inclusão de novos atributos ou a alteração da estrutura dos dados sem a necessidade de modificar um esquema predefinido.

Essa flexibilidade é especialmente útil em cenários nos quais os dados são altamente variáveis ou quando não se conhece antecipadamente a estrutura completa dos dados que serão armazenados. Os bancos de dados semi estruturados são amplamente utilizados em aplicações web, armazenamento de documentos, sistemas de gerenciamento de conteúdo e outras áreas em que a estrutura dos dados pode variar consideravelmente.

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

1. Flexibilidade: Os bancos de dados semiestruturados visam oferecer uma estrutura flexível para armazenar e organizar dados, permitindo a inclusão de informações com estruturas variáveis. Isso é particularmente útil em cenários em que a estrutura dos dados é desconhecida ou pode variar com o tempo [5].
2. Representação de dados complexos: Esses bancos de dados são projetados para lidar com dados não tabulares e complexos, como

documentos XML, JSON e texto livre. Eles fornecem uma maneira eficiente de armazenar e consultar dados que não se encaixam perfeitamente em um modelo de tabela relacional [5].

3. Recuperação eficiente de informações: Os bancos de dados semi estruturados permitem que as informações sejam recuperadas de maneira eficiente, mesmo em documentos com estruturas hierárquicas ou em formato de grafo. Eles oferecem recursos de consulta poderosos para buscar dados específicos, mesmo quando a estrutura dos dados não é rígida [5].
4. Integração com a Web: Uma das principais motivações para o desenvolvimento de bancos de dados semiestruturados foi a necessidade de lidar com dados provenientes da Web, onde a estrutura dos dados pode ser altamente variável. Esses bancos de dados foram projetados para facilitar o armazenamento, a recuperação e a integração de dados da Web de forma eficiente [5].

c) Vantagens da Tecnologia Pesquisada

1. Escalabilidade: Os bancos de dados semi estruturados são altamente escaláveis, o que significa que podem lidar com grandes volumes de dados de forma eficiente. Eles são projetados para lidar com a crescente quantidade de informações geradas nos dias atuais, bem como para suportar cargas de trabalho intensivas em leitura e gravação [1].
2. Flexibilidade de estrutura: Os bancos de dados semi estruturados permitem a inclusão de dados com estruturas variáveis. Isso significa que é mais fácil lidar com dados que não se encaixam perfeitamente em um modelo de tabela relacional, como documentos XML, JSON ou texto livre. Essa flexibilidade permite que os dados sejam adaptados e modificados facilmente à medida que as necessidades evoluem [1].
3. Recuperação de informações mais eficiente: Os bancos de dados semi estruturados oferecem recursos de consulta poderosos que permitem recuperar informações de maneira eficiente, mesmo quando a estrutura dos dados não é rígida. Isso facilita a busca e a análise de dados específicos, independentemente da variação na estrutura [1].

4. Adaptação a requisitos em constante mudança: Com a flexibilidade de estrutura dos bancos de dados semi estruturados, é mais fácil adaptar e modificar os dados à medida que os requisitos mudam. Isso permite que as organizações sejam mais ágeis e reajam rapidamente a novas demandas e oportunidades de negócios [1].

d) Desvantagens da Tecnologia Pesquisada

1. Ausência de restrições de integridade: Ao contrário dos bancos de dados estruturados, que possuem esquemas rígidos e restrições de integridade definidas, os bancos de dados semi estruturados geralmente não têm restrições pré-definidas. Isso pode levar a problemas de integridade dos dados, como a inclusão de informações inconsistentes ou incorretas [6].
2. Dificuldade na modelagem e design: A flexibilidade de estrutura dos bancos de dados semi estruturados pode tornar a modelagem e o design mais complexos. A ausência de esquemas rígidos requer uma cuidadosa consideração ao organizar e representar os dados, a fim de garantir a consistência e a eficiência das consultas [6].
3. Complexidade de consulta: Embora os bancos de dados semiestruturados ofereçam recursos poderosos de consulta, a complexidade da estrutura dos dados pode tornar as consultas mais complexas e menos intuitivas. A ausência de um esquema definido pode exigir conhecimento detalhado da estrutura dos dados para formular consultas precisas [6].
4. Performance variável: A flexibilidade de estrutura dos bancos de dados semi estruturados pode afetar a performance das consultas. Consultas complexas em dados semi estruturados podem exigir operações de busca e análise mais intensivas, resultando em tempos de resposta mais lentos em comparação com bancos de dados estruturados [6].
5. Gerenciamento de evolução dos dados: À medida que a estrutura dos dados semiestruturados evolui, pode ser necessário atualizar e adaptar os esquemas ou metadados associados. Isso pode exigir esforço adicional para o gerenciamento da evolução dos dados, garantindo que

as modificações sejam feitas de forma consistente e compatível [6].

e) Exemplo(s) de uso interessante(s) em empresas, projetos ou instituições

Exemplo de sucesso

Twitter guarda dados semi-estruturados. O *Twitter* é uma plataforma de mídia social que permite que os usuários publiquem tweets, que são mensagens curtas de até 280 caracteres. Cada tweet contém várias informações associadas a ele, como o conteúdo da mensagem, a data e a hora de publicação, o nome de usuário do autor, as hashtags usadas e outras metadados relevantes [2].

Embora os tweets sejam mensagens de texto relativamente curtas, o *Twitter* também permite que os usuários incluam elementos multimídia, como imagens, vídeos e links, em seus tweets. Esses elementos multimídia também são associados aos tweets e fornecem mais contexto e informações [2].

Embora os tweets sejam em sua maioria texto livre, o *Twitter* oferece suporte a algumas estruturas básicas, como hashtags, que são usadas para categorizar e agrupar tweets com base em tópicos específicos. Além disso, o *Twitter* usa tags e metadados internos para rastrear informações como menções a outros usuários, retweets e curtidas.

No entanto, é importante observar que os dados do *Twitter* podem ser considerados semi-estruturados porque nem todos os campos são obrigatórios e o conteúdo dos tweets pode variar significativamente. Isso contrasta com os dados estruturados, que possuem um formato fixo e campos bem definidos. Os dados semi-estruturados têm alguma forma de organização, mas também podem incluir elementos não estruturados.

Exemplo de insucesso

Um exemplo de insucesso envolvendo o uso de bancos de dados

semiestruturados ocorreu com o sistema de saúde do Reino Unido, especificamente com o projeto do sistema de registros eletrônicos de saúde conhecido como *National Programme for IT* (NPfIT). Esse programa, iniciado em 2002, visava centralizar e interconectar os registros médicos de pacientes em todo o país [3].

O NPfIT utilizou um banco de dados semiestruturado baseado em XML para armazenar os registros médicos. No entanto, o projeto enfrentou várias desvantagens relacionadas ao uso desse tipo de banco de dados [4]:

1. Ausência de esquema rígido: A falta de um esquema de dados rígido no banco de dados semiestruturado dificultou a garantia de consistência e qualidade dos dados. Isso resultou em problemas de integridade e precisão das informações dos pacientes.
2. Desempenho inferior: O banco de dados semiestruturado enfrentou dificuldades de desempenho, especialmente quando lidava com grandes volumes de dados e consultas complexas. Isso resultou em atrasos e tempos de resposta lentos, impactando negativamente a produtividade dos profissionais de saúde.
3. Esses desafios resultaram em atrasos significativos no projeto NPfIT e, eventualmente, no seu cancelamento em 2011. O projeto foi amplamente criticado por seus problemas técnicos, custos excessivos e falta de resultados tangíveis. O uso de um banco de dados semiestruturado contribuiu para as dificuldades encontradas no gerenciamento e na utilização eficaz dos registros eletrônicos de saúde.
4. É importante ressaltar que esses problemas não são exclusivos dos bancos de dados semi estruturados, mas sim desafios que podem surgir ao lidar com sistemas complexos de TI em grande escala. No entanto, o exemplo ilustra como as desvantagens dos bancos de dados semi estruturados podem afetar negativamente um projeto ou iniciativa, especialmente quando a integridade e a acessibilidade dos dados são fundamentais, como é o caso dos registros de saúde.

g) Base de Datos

Diagrama Entidade-Relacionamento

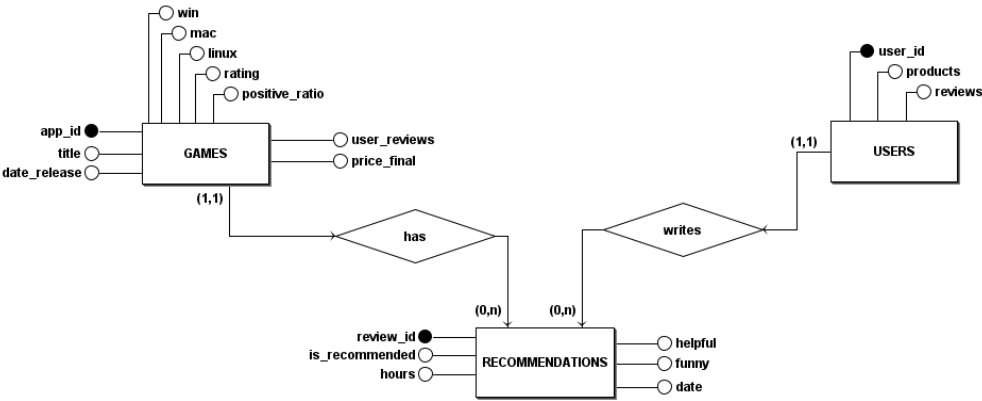
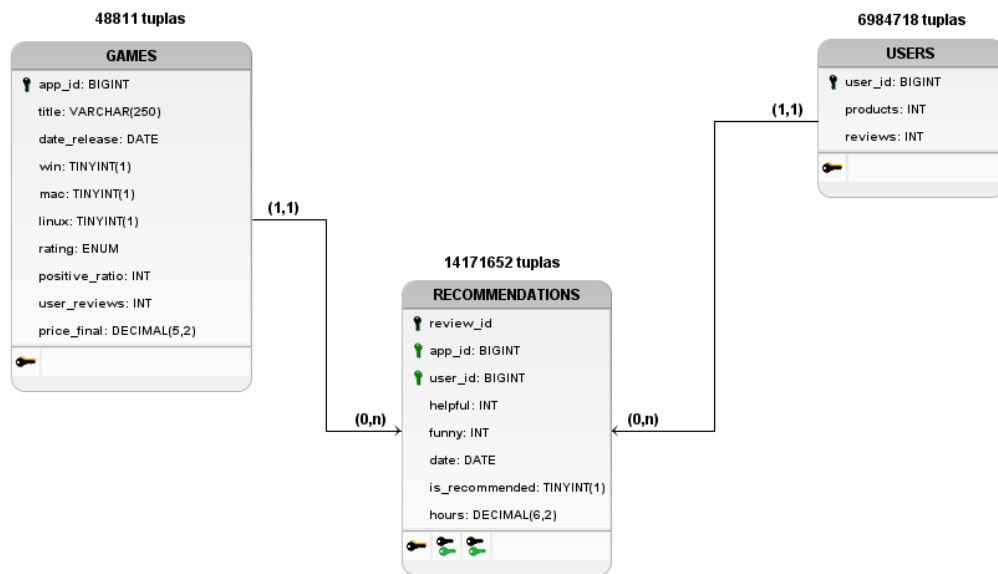


Diagrama Lógico de Datos



Endereço virtual da base de dados:

<https://www.kaggle.com/datasets/antonkozyriev/game-recommendations-on-steam?resource=download>

f) Bibliografias Pesquisadas

[1] ABITEBOUL, Serge; BUNEMAN, Peter; SUCIU, Dan; Data on the Web: From Relations to Semistructured Data and XML. Morgan Kaufmann, 2014.

[2] DATADITION. Is Twitter structured data or unstructured data? Disponível em:

<https://datadition.com/is-twitter-structured-data-or-unstructured-data/#:~:text=So%20presented%20broadly%2C%20%E2%80%9CTwitter%20data,is%20unstructured%20or%20structured%20data>. Acesso em: 10 jun. 2023

[3] COMPUTER WEEKLY. Six reasons why the NHS National Programme for IT failed. Disponível em:

<https://www.computerweekly.com/opinion/Six-reasons-why-the-NHS-National-Programme-for-IT-failed>. Acesso em: 10 jun. 2023.

[4] JUSTINIA, Taghreed. The UK's National Programme for IT: Why was it

dismantled? 2017. Health Serv Manage Res. Disponível em:
<https://pubmed.ncbi.nlm.nih.gov/28166675/#:~:text=The%20initiative%20was%20not%20trusted,officially%20dismantled%20in%20September%202011..>
Acesso em: 10 jun. 2023.

[5] AMBIKA, P.; Advances in Computers, v. 117, p. 321-338, 2020.

[6] GeeksforGeeks. What is Semi-Structured Data?. GeeksforGeeks. Disponível em: <https://www.geeksforgeeks.org/what-is-semi-structured-data/>. Acesso em: 10 jun. 2023.