

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

**Tecnologias de Banco de Dados (TI-BD)**

**Bancos de Dados Semiestruturados e XML**

**Leonardo dos Santos Silva Barreiros - 15/0135521**

Brasília, DF

2019

## Introdução

Será abordado por meio deste trabalho de pesquisa a respeito de Bancos de Dados Semi estruturados e XML, a fim de sintetizar quais os principais pontos positivos e negativos, como utilizar esta tecnologia e seus objetivos.

## Desenvolvimento

Antes de começar a abordar a respeito de banco de dados e XML devemos saber alguns conceitos que são importantes para o entendimento de como funciona esta tecnologia.

**Dados semiestruturados:** São uma forma de dados que apresentam uma representação estrutural heterogênea, não sendo completamente não-estruturados nem estritamente tipados.

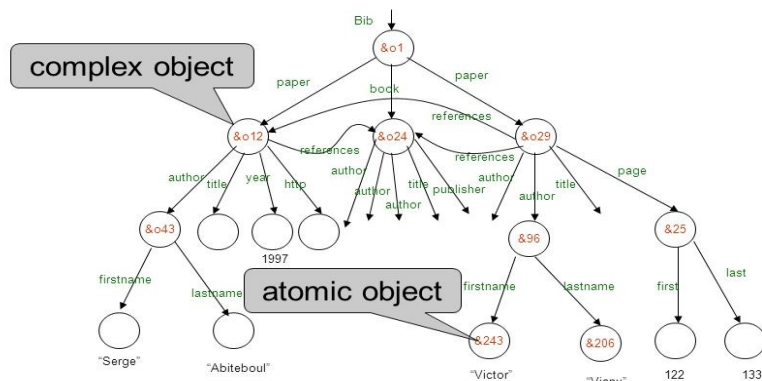


Retirado de: [https://universidadetecnologia.com.br/dados-estruturados-e-nao-estruturados/novo\\_baner\\_dados/](https://universidadetecnologia.com.br/dados-estruturados-e-nao-estruturados/novo_baner_dados/)

**Estrutura heterogênea:** São estruturas de dados que permitem a construção de estruturas mais complexas que os tipos de dados primitivos. Exemplo: int, float, char ...

Uma estrutura pode ser classificada em simples e complexa. Uma estrutura simples é atômica, ou seja, indivisível. Logo uma estrutura considerada complexa é um conjunto de estruturas simples que formam uma função.

## Modelo de dados semi-estruturado



Object Exchange Model (OEM)

Retirado de: <https://slideplayer.com.br/slide/5320401/> - slide 14

**Linguagens de marcação:** São utilizadas para definir formatos, seria uma maneira de exibição ou padrões aplicados em um documento. Servindo basicamente para definir um determinado conteúdo. Desta forma são utilizadas tag's ou marcador para que seja atribuído algum significado. Uma linguagem de marcação bem conhecida é o HTML. Os símbolos que utilizados para representação de marcação são: '<' '>' '/', '?' ... sendo estes os mais conhecidos e convencionais. Entretanto a linguagem que define seu tipo de marcação.

**XML:** Extensible Markup Language – conhecido como XML, é um formato de texto simples e flexível que vem do SGML(linguagem de marcação), que foi projetado para cobrir desafios de publicações eletrônicas em larga escala. Atualmente tem desempenhado papel, na maioria dos casos, entre sistemas e banco de dados, na web.

Tendo em vista os conceitos acima, podemos definir que um banco de dados semiestruturado com uso de XML é um tipo de forma ou gerenciamento de dados que não está totalmente, de acordo com as estruturas formais dos modelos dos SGBD's que conhecemos. Entretanto, devido ao XML ser pragmático no processo de integração entre bancos faz-se eficiente o uso, pois ainda que seja escrito um arquivo XML por um banco, não se torna tão custoso os impactos para que seja feita a leitura destes dados por um outro banco

qualquer.

Este exemplo demonstra a sintaxe flexível do XML sendo usada para descrever uma receita de pão:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<receita nome="pão" tempo_de_preparo="5 minutos" tempo_de_cozimento="1
hora">
  <titulo>Pão simples</titulo>
  <ingredientes>
    <ingrediente quantidade="3" unidade="xícaras">Farinha de
Trigo</ingrediente>
    <ingrediente quantidade="7" unidade="gramas">Fermento</ingrediente>
    <ingrediente quantidade="1.5" unidade="xícaras"
estado="morna">Água</ingrediente>
    <ingrediente quantidade="1" unidade="colheres de chá">Sal</ingrediente>
  </ingredientes>
  <instrucoes>
    <passo>Misture todos os ingredientes, e dissolva bem.</passo>
    <passo>Cubra com um pano e deixe por uma hora em um local
morno.</passo>
    <passo>Misture novamente, coloque numa bandeja e asse num
forno.</passo>
  </instrucoes>
</receita>
```

Retirado do site WikiPedia: <https://pt.wikipedia.org/wiki/XML>

Note que no exemplo acima temos uma grande semelhança com HTML. A tag principal <receita> tem como nomenclatura de *XML Schema*, ou seja, de acordo com a proposta da W3C *World Wild Web Consortium*, esta é para descrever a estrutura de um documento XML.

Desta forma podemos escrever um arquivo em XML, que poderá ser lido por quaisquer banco.

Observe que as tags de marcação não são nativas de nenhuma estrutura conhecida. Portanto, a partir da criação dessa estruturas temos o que chamamos de dados heterogêneos, mas estes dados são definidos em uma espécie de *struct* (como é conhecida na linguagem C) ou *library*. Dentro de cada uma destas tags existem funcionalidades determinadas para atribuir valor e significado para um determinado dado em que nela for aplicada.

Neste contexto de criação de dados em XML temos linguagens de consulta que tem como principal objetivo, das linguagens para dados semi-

estruturados, é consultar conjuntos de documentos como se fossem um BD, permitindo consultas mais eficazes e eficientes de dados, tendo em vista, que pode-se realizar um script convencional de um SGBD para carregamento dos dados desejados. Porém faz-se necessário atender alguns requisitos:

- Habilidade de executar consultas sem conhecimento do esquema ou sobre o próprio esquema, se este existir;
- Capacidade de lidar com a heterogeneidade de tipos de atributos, que podem ocorrer em objetos semanticamente iguais;
- Utilização de operadores de consulta de linguagens tradicionais de BDs, tais como seleção e projeção (passíveis de otimização);
- Navegação no estilo hipertexto e busca por padrão;
- Consultas temporais;
- Uso de uma abordagem baseada em objetos, uma vez que dados semi-estruturados são complexos;
- Uso de coerção (critérios de conversão de valores), em função da heterogeneidade de tipo e de estrutura dos dados semi-estruturados;
- Suporte à especificação de expressões de caminho (path expressions), a fim de se navegar pela estrutura de dados semi-estruturados e documentos;
- Busca baseada em estrutura, que permite restringir o escopo da consulta àqueles elementos que possuem uma determinada estrutura hierárquica implícita ou explícita.

Uma das grandes vantagens de se usar um banco com XML e para seu modelo relacional foi o fato de os dados estarem de acordo com um esquema e seguirem um conjunto de restrições de integridade. Exemplo para este caso: O programa só processaria esses documentos se eles estivessem de acordo com o esquema definido.

As desvantagens em geral se restringem às aplicações que não demandam maior complexidade, tais como vetores, listas associativas (chave-valor) e informações relativas a configuração, em que o bom senso estabelece a melhor escolha (entre o XML ou um formato menos popular).

Um exemplo de empresa que faz o uso do banco de dados semi-estruturado com XML é o Ministerio da cidadania na parte de gestão de levantamento de processos através da plataforma Redmine, onde esse sistema gera um banco com estruturas devidas para alimentar outra plataforma conhecida como Citesmart que faz utilização desses dados para preenchimento de planilha que ao final é fechado processos com dados especificos e um Identificador único daquele processo.

## **Conclusão**

A integração de dados é um assunto que desperta grande interesse não só da comunidade acadêmica, mas também de diversos outros setores. A necessidade das organizações por otimização em questões de lucratividade e eficiência dos resultados exige o uso de tecnologia de ponta para a diminuição de custos, a agilidade dos seus processos facilitação de leitura de dados, a fim de reduzir fronteiras. Para isso, uma forma de aperfeiçoar seus processos é a utilização da XML como mediador e integração de sistemas e/ou de dados. Entretanto, dependendo da magnitude do sistema pode-se tornar custoso e complexo o uso de bancos semi-estruturados com uso de XML, tornando mais viável o uso de bancos com padrões determinados, tanto para leitura de dados quanto para determinação dos dados e o uso do XML sendo feito para alimentação de sistemas como tem sido feito atualmente por maior parte dos arquitetos de sistemas de banco de dados mais convencionais.

## Referências

- Artigo SQL Magazine 8 DevMedia, Eduardo. A tecnologia XML e sua integração com Banco de Dados. Disponível em: <https://www.devmedia.com.br/artigo-sql-magazine-8-a-tecnologia-xml-e-sua-integracao-com-banco-de-dados/5147>. Acesso em: setembro de 2019.
- Artigo sobre dados semi-estruturados Universidade de São Paulo, Dados Semi-Estruturados. Disponível em: <https://www.ime.usp.br/~jef/semi-estruturado.pdf>. Acesso em: setembro de 2019
- Artigo linguagem de marcação Portal Educação, Linguagem de Marcação, Disponível em: <https://www.portaleducacao.com.br/conteudo/artigos/informatica/linguagem-de-marcacao/31639> . Acesso em: setembro de 2019.
- W3C – About XML About XML, W3C, Extensible Markup Language, Disponível em: <https://www.w3.org/XML/> . Acesso em: setembro de 2019.
- SGML TechTarget, SGML, Disponível em: <https://whatis.techtarget.com/definition/SGML-Standard-Generalized-Markup-Language>. Acesso em: setembro de 2019
- Livro Sistemas de Bancos de Dados, 6º edição. Editora Person Brasil (2002) – 808 páginas, Shamkant B. Navathe e Ramez Elmasri.