

SISTEMAS DE BANCO DE DADOS 2

AULA 15

Conhecendo melhor os Dados

Vandor Roberto Vilardi Rissoli



APRESENTAÇÃO

- Conhecendo mais os Dados
- Processar volume de dados
 - Tipos de dados
- Modelagem e representação
- Referências



CARACTERÍSTICAS DOS DADOS

A IMPORTÂNCIA DOS DADOS

- Dados sempre foram gerados desde tempos remotos a nossa realidade, porém o volume dessa geração nos tempos atuais é enorme.
- No início dessa década a **IBM** publicou que 90% de todos os dados do mundo foram gerados nos últimos 2 anos.



90% = últimos 2 anos

10% = desde o início



CARACTERÍSTICAS DOS DADOS

- Com tantos dados alguns desafios precisavam ser superados, destacando-se entre eles como:

- Armazenar tantos dados;
- Utilizar esses dados.



- O avanço de métodos e tecnologias coerentes a essa realidade têm sido fundamentais para tais superações

- *Data Warehouse*

- *Business Intelligence*

- *Big Data*

- *Data Mining*

- *Dashboard*

(visualização)



- entre outras.

CARACTERÍSTICAS DOS DADOS

HISTÓRIA DOS DADOS

- Dados são essenciais as pessoas e organizações;
- Guardados de diversas formas, mas evoluíram em seu armazenamento (físico => digital);

- Anteriormente, extrair informações e manter os dados organizados era uma tarefa muito **custosa**;



- O acesso aos dados dependia da **localização geográfica** dos arquivos que os armazenavam;

- Apenas armazenar não resolvia o problema, era necessário que os dados se relacionassem e pudessem ser usados pelas pessoas interessadas.



CARACTERÍSTICAS DOS DADOS



Fitas Magnética
Cartão Perfurado
Leitura de Dados
Sequencial



Edgar Frank Codd
propõe o modelo
de dados relacional.
E surge o termo
SGBDR.



Computador
Pessoal (PC)
Sistemas de Banco
de Dados
Linguagem SQL
Banco de Dados
Comerciais

1950 - 1960

1960 - 1970

1970 - 1980

1980

1980 - 1990



Discos Rígidos
Leitura Não
Sequencial
Modelo de Dados
Hierárquico
Modelo de Dados
em Rede

Dr. Peter Chen
propõe o modelo
Entidade-
Relacionamento



CARACTERÍSTICAS DOS DADOS

- Estima-se que diariamente são gerados 15 Petabytes de informações (redes sociais, dispositivos móveis, financeiros) em todo o mundo, provenientes de diversas plataformas e sistemas;



Facebook 10 e Twitter 7 Terabytes diários cada um;

Dados gerados nas **pesquisas astronômicas** armazenaram só em 2010 cerca de 140 Terabytes, tendo a expectativa desse volume de dados ser gerado a cada 5 dias com os novos **telescópios**.



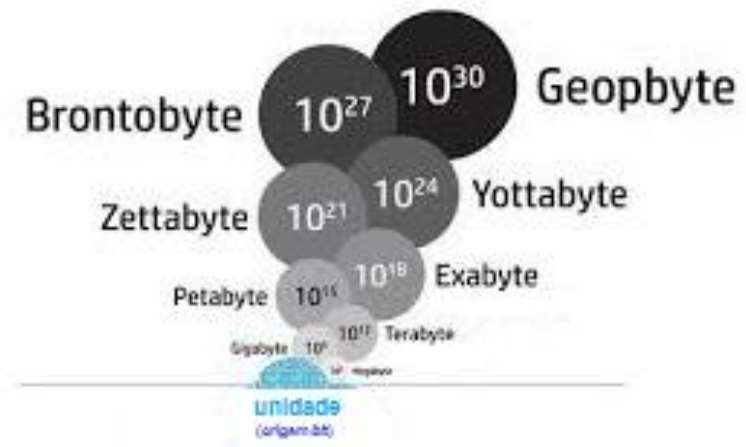
CARACTERÍSTICAS DOS DADOS

VOLUME DE DADOS

- No ano 2000, eram armazenados no mundo 800.000 Petabytes (PB)
- A expectativa da IBM para 2020 é a geração de 35 Zettabytes (ZB)

SISTEMA DE MEDIDAS EM ESPAÇO DE MÍDIAS ELETRÔNICAS

MEDIDA USUAL	Nº DE CARACTERES	BASE 2
Byte	1	2^0
Kilobyte (KB)	1.024	2^{10}
Megabyte (MB)	1.048.576	2^{20}
Gigabyte (GB)	1.073.741.824	2^{30}
Terabyte (TB)	1.099.511.627.776	2^{40}
Petabyte (PB)	1.125.899.906.842.624	2^{50}
Hexabyte (HG)	1.152.921.504.606.846.976	2^{60}



- Como processar esse volume enorme de dados?



<https://www.youtube.com/watch?v=hEFFCKxYbKM>

CARACTERÍSTICAS DOS DADOS

- Um exemplo seria a elaboração da primeira foto de um buraco negro obtida pelo ser humano.
 - Consórcio internacional de radiotelescópios flagra as cercanias do fenômeno no centro da galáxia Messier 87 (M87), distante 55 milhões de anos-luz do sistema solar;

Anel com coloração artificial mostrando a porção mais interna de um disco ultra-aquecido de gás (ordem de 100 milhões de graus).



Sítio virtual origem: <https://revistapesquisa.fapesp.br/2019/04/10/a-primeira-imagem-dos-arredores-de-um-buraco-negro/>

CARACTERÍSTICAS DOS DADOS

- Resultado do esforço de mais de 200 pesquisadores do projeto Telescópio Horizonte de Eventos (EHT), consórcio internacional que usou dados coletados por oito observatórios de radioastronomia espalhados dos Estados Unidos ao pólo Sul do planeta, programados para funcionar em rede como se fossem um único radiotelescópio com o tamanho aproximado da Terra;

Rede de radiotelescópios do Atacama Large Millimeter/Submillimeter Array (Alma), no Chile, que participou das observações realizadas em 2017.

NRAO/AUI/NSF



CARACTERÍSTICAS DOS DADOS

- As observações para produção dessa imagem são de cerca de 5 petabytes de dados (5 milhões de gigabytes ou algo próximo a 5 mil anos de música guardada em arquivos digitais), que tiveram de ser armazenados em centenas de discos rígidos e transportados de avião para dois centros de análise: o Observatório Haystack no MIT (EUA) e o Instituto Max Planck de Radioastronomia na Alemanha.



CARACTERÍSTICAS DOS DADOS

- Nesses centros, as equipes trabalharam por quase dois anos para correlacionar os dados dos diferentes observatórios e fazer as devidas correções;
- As características do buraco negro da M87, cuja massa, calculada agora, é 6,5 bilhões de vezes maior que a do Sol do nosso sistema solar;

A galáxia M87 é um dos jatos de partículas que emana da vizinhança de seu buraco negro, observados pelo telescópio espacial Hubble.

NASA, ESA, Hubble Heritage Team (STScI/AURA), P. Cote (Instituto de Astrofísica Herzberg) e E. Baltz (Universidade Stanford).



CARACTERÍSTICAS DOS DADOS

DADOS ESTRUTURADOS

- Dados organizados em blocos semânticos (relações ou tabelas);
- Dados de um mesmo grupo possuem as mesmas descrições (atributos);
- Descrições para todas as classes de um grupo possuem a mesma estrutura / formato (esquema);
- Dados mantidos em um SGBD são chamados de **Dados Estruturados** por manterem a mesma estrutura de representação (rígida), previamente projetada (esquema).

Nome	Idade	Departamento	Salário	Telefone
Alberto da Silva	25	Vendas	R\$ 850,00	555-1902
Antônio dos Santos	32	Administração	R\$ 1.200,00	555-1117
Fabiana Rossi	40	Administração	R\$ 2.000,00	555-8929
Horácio Almeida	31	Recursos Humanos	R\$ 1.350,00	555-8907
João Pereira	35	Vendas	R\$ 1.500,00	555-7814
Roberto Albuquerque	29	Administração	R\$ 1.200,00	555-8273
Sônia Pires	23	Vendas	R\$ 600,00	555-8664



CARACTERÍSTICAS DOS DADOS

DADOS SEMIESTRUTURADOS

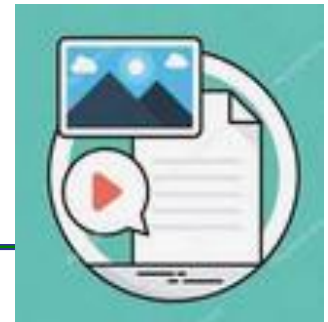
- Muitos dados atualmente **NÃO** são mantidos no Sistema Gerenciador de Banco de Dados - SGBD;
- Dados Web, por exemplo, apresentam uma organização bastante **HETEROGÊNEA**;
- A alta heterogeneidade **DIFICULTA AS CONSULTAS** a estes tipos de dados;
- Assim, estes dados são classificados como **Dados Semiestruturados**:
 - Não são estritamente “tipados”;
 - Não são complementarmente não-estruturados.



CARACTERÍSTICAS DOS DADOS

NÃO ESTRUTURADOS

- São os dados que NÃO possuem uma estrutura definida para armazenamento;
- Caracterizados, normalmente, por documentos textos, imagens, vídeos, etc.;
- Nem as estruturas são descritas implicitamente como nos semiestruturados;
- Grande maioria dos dados atuais na Web e nas empresas seguem este formato (estima-se que 80% de todos os dados armazenados no mundo).



CARACTERÍSTICAS DOS DADOS

- Comparação entre as características dos tipos

Dados Estruturados	Dados Semiestruturados	Dados Não Estruturados
Esquema pré-definido	Nem sempre há um esquema	Não há esquema
Estrutura regular	Estrutura irregular	Estrutura irregular
Estrutura independente dos dados	Estrutura embutida nos dados	Pode não ter estrutura alguma
Estrutura reduzida	Estrutura extensa (particularidades de cada dado, visto que cada um pode ter uma organização própria)	Estrutura extensa (particularidades de cada dado, visto que cada um pode ter uma organização própria)
Fracamente evolutiva	Fortemente evolutiva (estrutura modifica-se com frequência)	Fortemente evolutiva (estrutura modifica-se com frequência)
Prescritiva (esquemas fechados e restrições de integridade)	Estrutura descritiva	Estrutura descritiva
Distinção entre estrutura e dados é clara	Distinção entre estrutura e dados não é clara	Distinção entre estrutura e dados não é clara



CARACTERÍSTICAS DOS DADOS

- Como então trabalhar com diferentes tipos de dados em volumes significativos, enormes ou não?



CARACTERÍSTICAS DOS DADOS

- A realidade demanda de tal necessidade de armazenamento contínuo, com dados que representem a realidade de maneira fidedigna aos acontecimentos, desejos e agilidade;
- A manipulação desse universo de dados armazenados também necessita de agilidade para atender em tempo hábil as demandas existentes.
- O que seria a *Fast Data* diante dessa realidade?



<https://www.youtube.com/watch?v=2ZWIYCLBeB0>



Referência de Criação e Apoio ao Estudo

Material para Consulta e Apoio ao Conteúdo

- SILBERSCHATZ, A. e KORTH, H. F., Sistemas de Banco de Dados, tradução da 6ª edição
 - Capítulo 20
- WIKILIVROS – SQL/ Dados Estruturados, Semiestruturados e Não Estruturados
 - Site:
https://pt.wikibooks.org/wiki/SQL/Dados_Estruturados,_Semi-Estruturados_e_N%C3%A3o_Estruturados
- Universidade de Brasília (UnB Gama)
 - Site: <https://cae.ucb.br/conteudo/unbfga>
(escolha a disciplina **Labor. Banco Dados (SBD2)**
e a opção **Outras Tecnologias de Banco de Dados**)

