

Gestão da Tecnologia da Informação

Introdução a Big Data

Francisco José Tosi



EDUCAÇÃO
METODISTA

Introdução a Big Data

Aula 07 Tipos de Big Data

12/04

Francisco José Tosi



EDUCAÇÃO
METODISTA



Introdução a Big Data

Tipos de Big Data

Dados Estruturados



Dados Não Estruturados



EDUCAÇÃO
METODISTA

Introdução a Big Data

Tipos de Big Data

O Big Data trouxe dois novos fatores a serem considerados



Dados gerados a partir
de sensores,
smartphones e tablets



A criação de uma maneira
rentável de lidar com toda
essa quantidade de dados



EDUCAÇÃO
METODISTA

Introdução a Big Data

Tipos de Big Data

Análise em tempo REAL: quando existe a necessidade de análise em tempo real.

Ex.: Monitoramento de dados de tráfego.

Análise em lotes: análise de grande quantidade de dados para determinar um padrão necessário.

Ex.: Identificar padrão de compras de clientes pela internet.

Integrar dados: integrar múltiplas fontes de dados para obter soluções.

Ex.: Dados de redes sociais ou Tv com comportamento de compras das pessoas.

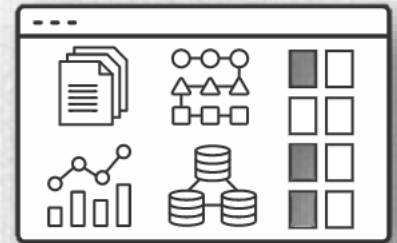
Introdução a Big Data

Dados Estruturados

Dados que possuem um formato e comprimento definidos

Exemplo: Números; Datas; grupo de palavras

Abrange 20% dos dados existentes



São armazenados em uma base de dados

Podem ser consultados utilizando uma linguagem estruturada de consulta SQL
(structured Query Language)

Introdução a Big Data

Dados Estruturados

Dados Estruturados: são aqueles que possuem algum padrão ou formato, que podem ser usados na sua leitura e extração dos dados.

Dados de bancos de dados, sistemas legados, arquivos texto.



Exemplos: csv, txt ou xml.



Introdução a Big Data

Dados Estruturados

Podem ser coletados:

Gestão de relacionamento com cliente (CRM – Customer Relationship Management)

Dados de sistema integrado de gestão empresarial (SIGE)

Dados financeiros



EDUCAÇÃO
METODISTA

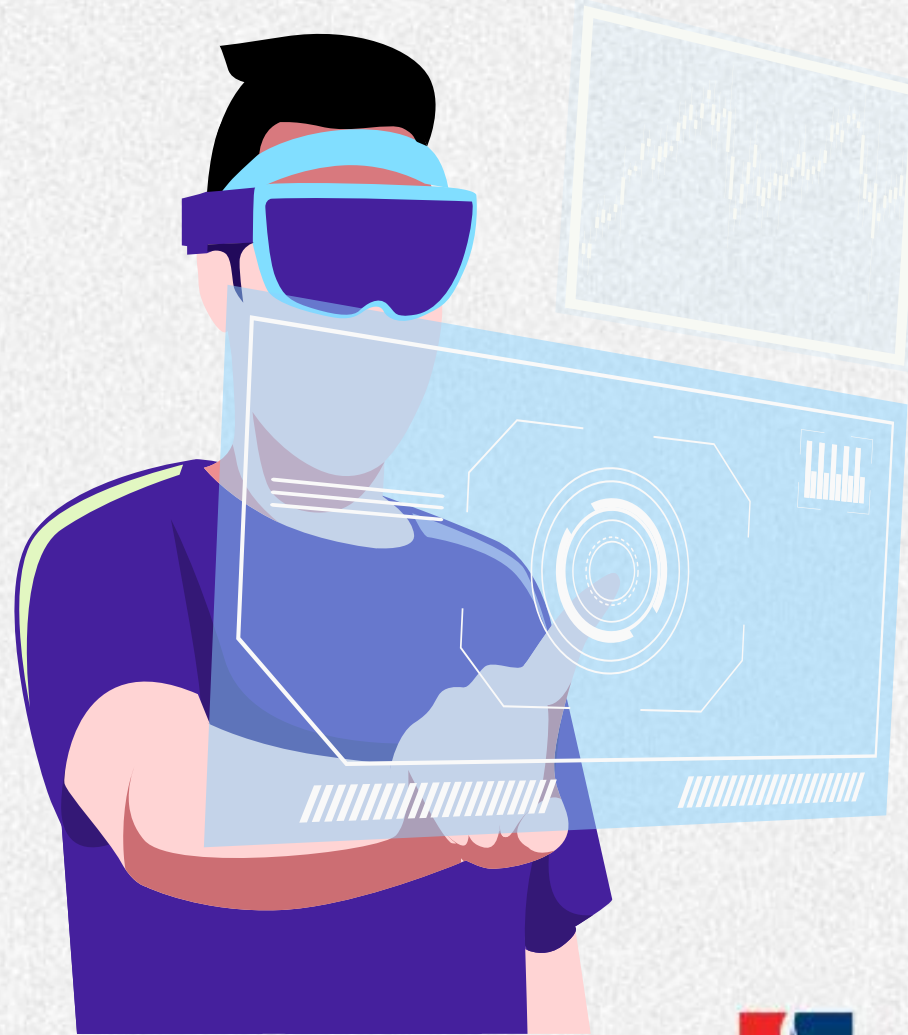
Introdução a Big Data

Dados Estruturados

Fontes são divididos em duas categorias:

Gerados por máquinas: são gerados por máquinas sem interferência humana

Gerados por humanos: são gerados por humanos, através de interações com computadores



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dados Estruturados

Dados gerados por máquinas:

Dados do sensor: etiquetas de radiofrequência, medidores inteligentes, dispositivos médicos dados de sistemas de posicionamento global (GPS)

Dados de web log: quando servidores, aplicativos, redes e outros operam, capturam todos tipos de dados sobre suas atividade que geram enormes volumes de dados que podem ser úteis

Dados de pontos de vendas: um caixa de supermercado passando o código de barras de produtos, todos os dados referentes ao produto são gerados. Pense em todos os produtos e informações obre os produtos.

Dados financeiros: sistemas financeiros são programáticos, operam com base em regras predefinidas que automatizam processos. Dados de negociações de ações.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dados Estruturados

Dados gerados por humanos:

Dados de entrada: todos os dados que uma pessoa pode colocar no computador, ex: nome, idade, renda, respostas de pesquisas com formulários restritos. Dados que são utilizados para entender o comportamento de um cliente

Dados de fluxo de clique: dados gerados nos cliques efetuados em links de website, são utilizados para analisar e determinar o comportamento de clientes e padrões de compras

Dados relacionados a jogos: cada movimento realizado em um jogo são gravados. Isso pode ser útil para entender como usuários finais se movem através de um portfólio de jogos



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dados Estruturados

O Papel da base de dados **relacionais** em Big Data

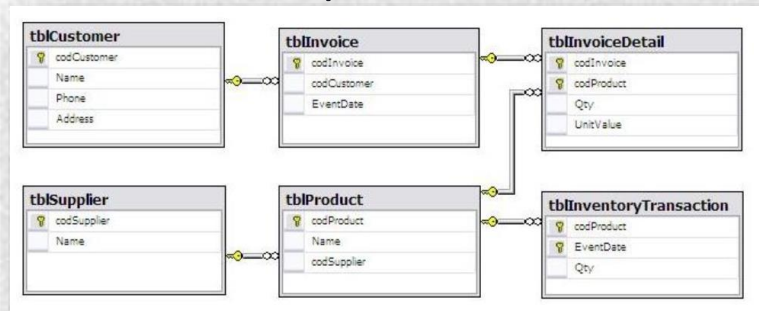


Persistência de dados: é como a base de dados retém versões dela mesma quando modificada.

O início do armazenamento de dados persistentes é o SISTEMA DE ADMINISTRAÇÃO DE DADOS RELACIONAIS - RDBMS

Os dados são armazenadas em uma tabela

Ela contém uma representação estrutural do que está dentro da base de dados - ESQUEMA



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dados Não Estruturados

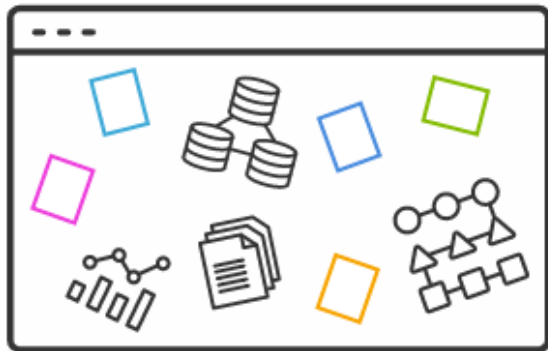
Não seguem formato específico

80% dos dados de empresas são desestruturados

Até recentemente a tecnologia não conseguia suportar a manutenção destes tipos de dados, exceto armazená-los ou analisá-los manualmente

Dados desestruturados estão por todas as partes

São gerados por máquinas e humanos



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dados Não Estruturados

Dados Não-Estruturados são aqueles que **não** possuem um formato padronizado para leitura -podem ser arquivos.



Word, Páginas de Internet/Intranet,
Vídeos, áudios, entre outros



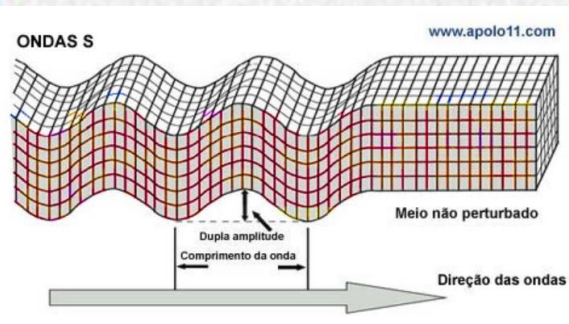
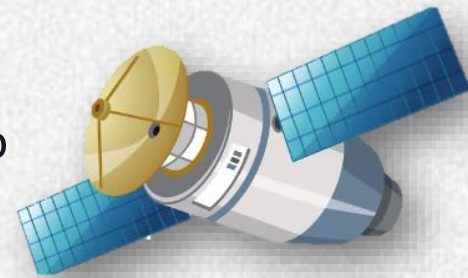
EDUCAÇÃO
METODISTA

Introdução a Big Data

Dados Não Estruturados

Dados gerados por máquinas:

Imagens de satélites: Informações de tempo ou que o governo captura em suas imagens de vigilância por satélite
Ex. Google Earth



Dados científicos: Imagens sísmicas, dados atmosféricos e físicos de altas energias



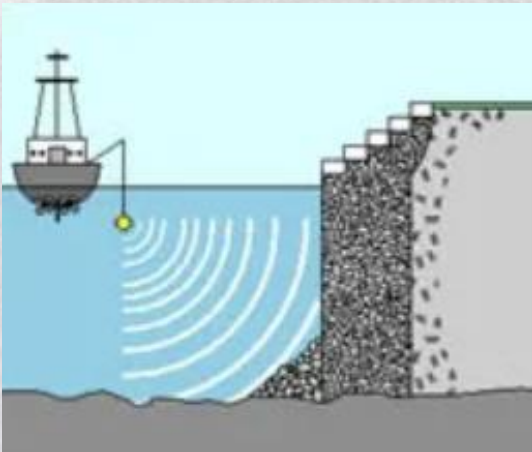
EDUCAÇÃO
METODISTA

Introdução a Big Data

Dados Não Estruturados

Dados gerados por máquinas:

Fotografias e vídeos: vídeo de segurança, vigilância e tráfego



Dados de radar ou sonar: Inclui perfis sísmicos veiculares , meteorológico e oceanográficos

Introdução a Big Data

Dados Não Estruturados

Dados gerados por humanos:

Texto próprio de empresas: Documentos, registros, resultados de pesquisas, e-mails.



Dados de radar mídia social: plataformas de mídias sociais, YouTube, Instagram



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dados Não Estruturados

Dados gerados por humanos:

Dados móveis: mensagem de texto e informações de localização.



Dados website: Sites com conteúdo desestruturados, YouTube, Flickr, Instagram

Introdução a Big Data

Dados Não Estruturados

Os dados desestruturados estão expandindo rapidamente

Podem ser utilizados analisar dados relevantes e transforma-los em informações estruturadas

Por exemplo:



Analisar mídia social para uso de uma grande quantidade de conversas de clientes

Dados de call center, e-mails, comentários escritos em uma pesquisa, documentos são analisados para entender o comportamento do cliente

Introdução a Big Data

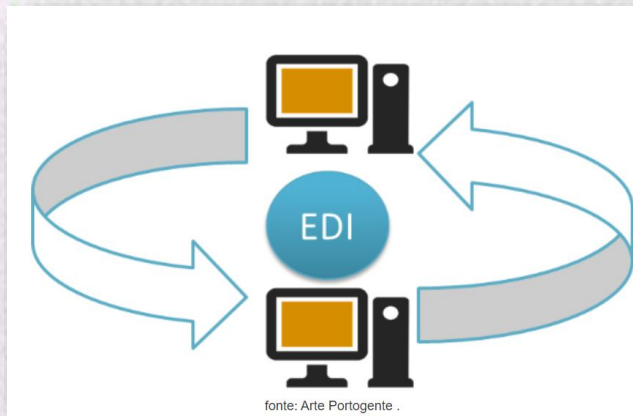
Dados semiestruturados

Ficam entre os estruturados e desestruturados

Não se conformam com o esquema fixo, uma estrutura

Mas podem ser auto descritos e podem ter partes de etiquetas/valores simples

Exemplo:



EDI

SWIFT

XML

Introdução a Big Data

Para gerenciamento e armazenamento de dados não estruturados pode ser utilizado CMS – Content Management System

Segundo Association for Information and Image Management – AIIM

www.aiim.org

ECM Enterprise Content Management é uma ferramenta que compreende as estratégias, métodos e ferramentas para Capturar, administrar, armazenar, preservar e entregar conteúdo e documentos relacionados a processo organizacionais



Introdução a Big Data

CMS para Administrar Big Data



Alguém postou no Twitter que está procurando algum produto

Uma empresa realiza uma pesquisa CMS e responde para a pessoa um local e onde pode encontrar o produto

Esse tipo de interação pode acontecer em tempo real

Na realidade devemos utilizar uma abordagem híbrida para resolver os problemas de Big Data

Introdução a Big Data

Administrando tipos de dados

	Carga	Fluxo	Consulta completa
Estruturados	Hadoop	Chave/valor	RDBMS
Desestruturados	Documento	Gráfica especial	Colunar
Ambos	Hibrido	Hibrida	Hibrida



Introdução a Big Data

Integrando tipo de dados em um ambiente Big Data

Não é necessário possuir todos os dados que necessita

Exemplo: podemos utilizar dados de mídia social, e de um satélite



Os dados podem estar vindos de sistemas internos ou externos

Alguns dados podem ter sido armazenado em momentos diferentes

Os dados não precisam chegar em tempo real



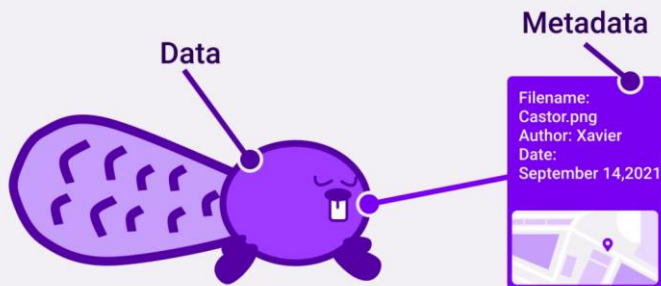
EDUCAÇÃO
METODISTA

Introdução a Big Data

Integrando tipo de dados em um ambiente Big Data

Conectores: Permitem extrair dados de várias fontes de Big Data.

ex. um conector do Twitter ou de um Facebook, ou integrar seu armazém de dados a uma fonte Big Data fora de seu local



Metadados: é o componente principal para integrar todos os dados, são as definições, mapeamentos e outras características usadas para descrever, como encontrar, acessar e utilizar os componentes de dados.

Ex. informações sobre a conta, número, descrição, tipo de dados, etc.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Atividade 3

Realizar a leitura individual do artigo:

Data Warehouse / Armazém de Dados: O que é, conceito e definição

Disponível em:

<https://cetax.com.br/o-que-e-data-warehouse/>

Em grupo de, fazer um debate com o que foi entendido sobre o artigo.

Em grupo produzir um resumo do artigo e disponibilizar no Moodle.

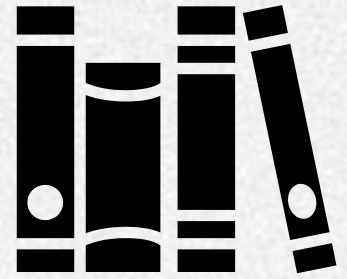
O Assunto abordado neste artigo será pedido em prova



EDUCAÇÃO
METODISTA



Introdução a Big Data



Basso, Douglas Eduardo; Big Data [recurso eletrônico] / Douglas Eduardo Basso, Curitiba: Contentus, 2020.

Informática; O grande livro do Big Data: Um guia prático para tirar seu primeiro projeto de Big Data do papel

I2AI, a conect al Word; Desmistificando Machine Learning, acessado de <https://www.i2ai.org/>

Intel IT Center, Guia de planejamento: Introdução à Big Data, como avançar com uma implantação bem-sucedida; 2014

Hurwitz, Judith; Big Data para leigos / Judith Hurwitz, Alan Nugent, Dr. Fern Halper, Marcia Kufman – Rio de Janeiro : Alta Books, 2015

Taurion, Cezar; Big Data / Cezar Taurion – Rio de Janeiro: Brasport, 2013



EDUCAÇÃO
METODISTA