

Gestão da Tecnologia da Informação

Introdução a Big Data

Francisco José Tosi



EDUCAÇÃO
METODISTA

Introdução a Big Data

Aula 03

Compreendendo o Big Data

14/03

Francisco José Tosi



EDUCAÇÃO
METODISTA

Introdução a Big Data

Artigo - O que é um banco de dados colunar

<https://kondado.com.br/blog/blog/2019/10/31/o-que-e-um-banco-de-dados-colunar/>

banco de dados tradicionais mais conhecidos, armazenam informações lineares

São criados índices, onde facilitam a busca das informações

Porem para saber informações de várias item do banco deverá ser acessado todos os itens, e como, em um baco de dados as informações estão no HD e não na memória deve ser uma operação de leitura para cada item



Introdução a Big Data

Artigo - O que é um banco de dados colunar

<https://kondado.com.br/blog/blog/2019/10/31/o-que-e-um-banco-de-dados-colunar/>

1	2	3	João	Carlos	Kondado	Pereira	Gonçalves	Inteligência	32	41	13
---	---	---	------	--------	---------	---------	-----------	--------------	----	----	----

1	João	Pereira	32	2	Carlos	Gonçalves	41	3	Kondado	Inteligência	13
---	------	---------	----	---	--------	-----------	----	---	---------	--------------	----

A leitura é simplificada pois é realizada somente nas colunas que serão utilizadas



Introdução a Big Data

Artigo - O que é um banco de dados colunar

<https://kondado.com.br/blog/blog/2019/10/31/o-que-e-um-banco-de-dados-colunar/>

“Os bancos de dados colunares disponíveis no mercado desenvolveram outras otimizações para melhoria de performance, o que os torna ideais para times de analytics que precisam lidar com grandes bases de dados – principalmente otimização de cruzamentos entre tabelas.”



EDUCAÇÃO
METODISTA

Introdução a Big Data

Porque "Big Data" é importante?

Permite que as organizações recolham, armazenem, administrem e manipulem grandes quantidades de dados na velocidade certa, no tempo certo, para conseguir conhecimentos corretos.

A tecnologia permite pegar cada pequeno detalhe da empresa e, por meio da coleta, processamento e análise de dados, encontrar possibilidades de melhoria.



Big Data trouxe avanços para diversas áreas, em especial para os setores de marketing e vendas, ajudando as organizações a reduzir custos, aumentar a produtividade e tomar decisões mais embasadas e inteligentes.



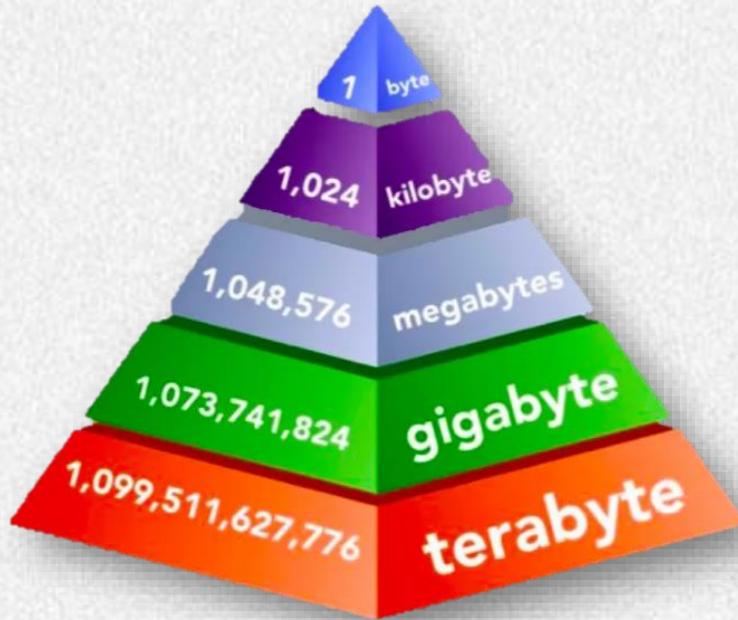
EDUCAÇÃO
METODISTA

Introdução a Big Data

O que é exatamente "Big Data"?

Big Data não é um grande Data Warehouse?

Big Data não é um BI em um data set de terabytes de dados?



Podemos dizer que sim

Mas é muito mais que isso



EDUCAÇÃO
METODISTA

Introdução a Big Data

O que é exatamente "Big Data"?

Big Data Não é uma tecnologia

Big Data é uma combinação de tecnologias novas e antigas que ajudam as empresas conseguirem ideias viáveis

É a capacidade de administrar um volume enorme de dados diferentes, na velocidade certa, dentro do prazo.



Introdução a Big Data

O que é exatamente "Big Data"?

Refere-se a grandes conjuntos de dados que estão em ordens de magnitudes maiores (**volume**),

incluindo dados estruturados, semiestruturados e não estruturados (**variedade**),

e chegando mais rapidamente (**velocidade**)

Essa avalanche de dados é gerada por dispositivos conectados, PCs e smartphones a sensores, como leitores de RFID e câmeras de trânsito

Além disso, é heterogênea e vem em muitos formatos, incluindo texto, documentos, imagens, vídeos, weblogs, transações e muito mais



Introdução a Big Data

Evolução da gestão dos dados

A evolução dos dados parte necessariamente de etapas anteriores.

De novos estágios ou onda de dados.

A gestão de dados deve incluir avanços em hardware, armazenamento, rede e modelos de computação como virtualização e computação em nuvem.



Introdução a Big Data

Evolução da gestão dos dados

Na necessidade de resolver problemas específicos do modelo de dados nasce uma onda de gestão de dados.

Cada uma das ondas ou fases evoluiu devido a causa e efeito, ou seja quando uma nova tecnologia chegava ao mercado, exigia a descoberta de novas abordagens.

Quando surgiu a base de dados relacional, precisou um conjunto de ferramentas que permitisse análise do relacionamento entre os elementos de dados.

Quando iniciou o armazenamento de dados desestruturados, precisou de ferramentas de análise baseada em linguagem natural.



EDUCAÇÃO
METODISTA

Introdução a Big Data

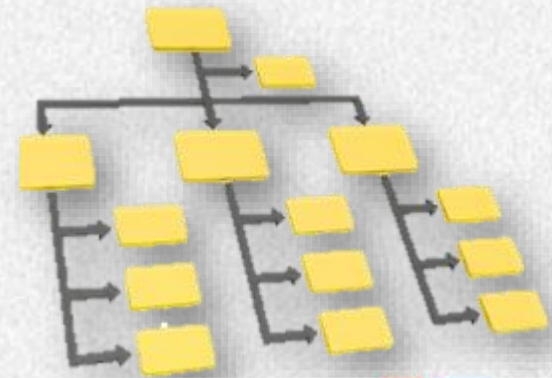
Evolução da gestão dos dados

Onda 1 – Criando estrutura de dados administráveis

Na década de 60, os dados eram armazenados em arquivos simples sem criação de estrutura.

Para obter entendimento detalhado sobre cliente, era necessário criar modelos de programação detalhados.

Em meados de 1970 foi criado o modelo de dados relacional e o sistema de administração de base de dados relacionais.



Com modelo de dado relacional era imposto um método que melhora o desempenho.

Introdução a Big Data

Evolução da gestão dos dados

Onda 1 – Criando estrutura de dados administráveis

O modelo ajudava as empresas a organizar melhor seus dados e serem capazes de comparar transações de uma geografia para outra.



Porém armazenar este volume de dados era caro e o acesso era lento.

Existia muitas duplicações de dados e era difícil medir o valor dos dados.

Gerando a necessidade de um criar um conjunto de tecnologia para apoiar o modelo relacional.



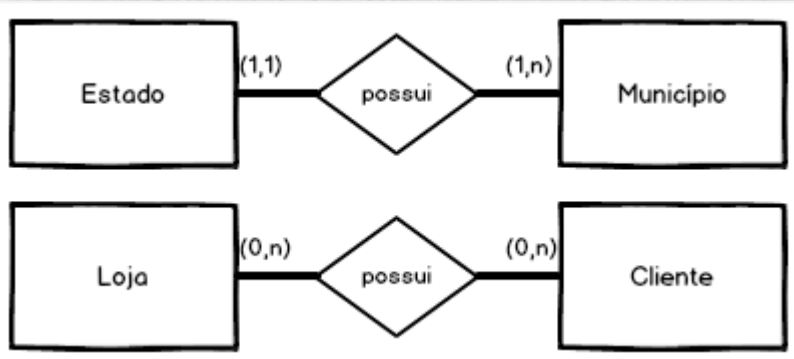
EDUCAÇÃO
METODISTA

Introdução a Big Data

Evolução da gestão dos dados

Onda 1 – Criando estrutura de dados administráveis

Surgiu o modelo Entidade e Relacionamento (ER) que adicionava uma abstração a mais para aumentar a usabilidade dos dados.



Os item eram definidos independente do seu uso.

Era possível criar relacionamento entre as fontes sem gerar códigos complexos.

Também foi possível a criação de modelos mais complexos



Introdução a Big Data

Evolução da gestão dos dados

Onda 1 – Criando estrutura de dados administráveis



Com o aumento do volume de dados que as empresas necessitam administrar, surge o armazém de dados.

Os armazéns de dados permitiu que as organizações de TI selecionasse um subconjunto de dados que estavam sendo armazenados.

Os armazéns ajudava as empresas a trabalhar com quantidade de dados estruturados maiores.

Com os armazéns era possível guardar dados de anos anteriores para a compreensão do desempenho empresarial.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Evolução da gestão dos dados

Onda 1 – Criando estrutura de dados administráveis



A necessidade das empresas passou a ser administrar dado transacionais em massa.

Para grandes volumes de dados estruturados e semiestruturados, os armazéns não foram capazes de evoluir o suficiente para atender as necessidades.

Os armazéns são alimentados em lotes semanais ou diários.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Evolução da gestão dos dados

Onda 1 – Criando estrutura de dados administráveis



É criado o sistema de administração de base de dados orientada a objeto – ODBMS (*Operacional Database Management Systems*).

Este sistema tem capacidade de armazenar grande objeto binário (BOGs – *binary large objects*).

Esta base tem uma linguagem de programação e uma estrutura para elementos de dados fácil de manipular vários objetos de dados.

As bases de dados orientada a objetos introduziram um novo nível de inovação que ajudou a conduzir para a segunda onda de gestão de dados.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Evolução da gestão dos dados

Onda 2 – Administração WEB e de conteúdo



Com o surgimento da web em 1990, as empresas partiram para um modelo mais unificado.

Mudança do modelo desconectado para um mais unificado.

União dos elementos da web, áudio e vídeo com uma plataforma incorporando a gestão de processos de negócios, controle de versão, reconhecimento da informação e administração de texto



EDUCAÇÃO
METODISTA

Introdução a Big Data

Evolução da gestão dos dados

Onda 3 – Administração o Big Data

Big data é o novo modelo e evolução na jornada de gestão de dados.

Foi construído em cima das práticas da gestão de dados das últimas ondas.

O custo dos ciclos de computação e armazenamento permitiu o armazenamento de mais informações.

Com o Big Data ficou possível virtualizar dados para serem guardados eficientemente baseados em nuvem.



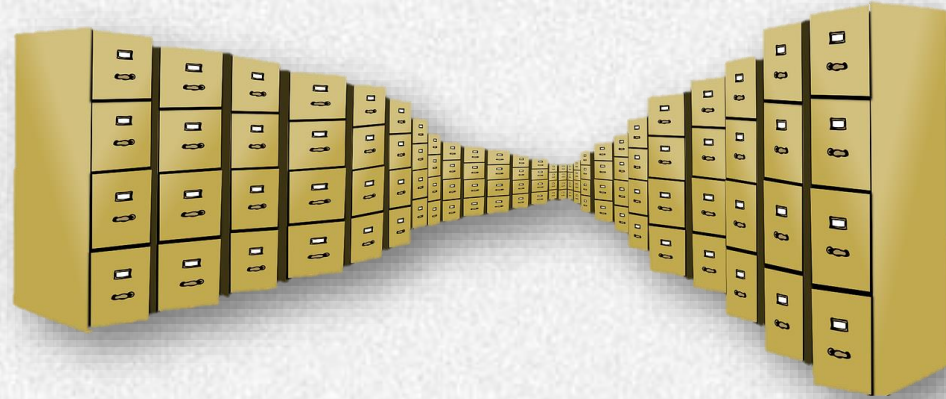
EDUCAÇÃO
METODISTA

Introdução a Big Data

Evolução da gestão dos dados

Onda 3 – Administração o Big Data

Com o Big Data é possível analisar petabytes (o equivalente a 20 milhões de gabinetes com 4 gavetas cheia de arquivos de texto).



13,3 anos de conteúdo HDTV com desempenho aceitável.

Não exclusivamente em negócios.



Pode ajudar pesquisa, atividades de governo

Introdução a Big Data

Evolução da gestão dos dados

Armazém de dados

Funcionando a partir de vários bancos de dados, um armazém de dados é organizado em seções à medida que recebe dados de bancos de dados operacionais e sistemas transacionais.



Os dados passam por vários processos de depuração à medida que entram em um armazém de dados, limpando e consolidando efetivamente os dados, tornando-os alguns dos dados mais confiáveis para uso da empresa.

Com o objetivo de análise aprofundada e criação de metadados, os data warehouses fazem o trabalho de transformar os dados de seu estado bruto em esquemas de várias camadas.

Os analistas podem avaliar e interpretar em cada camada da infraestrutura de dados para obter insights valiosos de negócios por meio de acesso autorizado.

Introdução a Big Data

Definindo o Big Data



O Big Data é uma combinação de tecnologias novas e antigas.

O Big Data é a capacidade de administrar um volume enorme de dados diferentes.

Na velocidade certa no prazo certo a fim de permitir análises e reações em tempo real.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Definindo o Big Data

Para iniciar os dados devem ser capturados, organizados, integrados.

Após a implementação os dados devem ser analisados com base no problema a ser tratado.

A gestão entra em ação com base nos resultados das análises.

O processo deve iniciar com o problema a ser resolvido, o que determinará o tipo de dados que será necessário e a arquitetura que deve ter



EDUCAÇÃO
METODISTA

Introdução a Big Data

Definindo o Big Data

Exemplo:



A Amazon pode indicar um livro com base nas compras anteriores do cliente.

Ou dar um cupom de desconto para a próxima compra de um produto relacionado a outro que foi comprado.



Introdução a Big Data

Definindo o Big Data

Estabelecendo a base de arquitetura

Definir os requisitos funcionais e dar suporte ao desempenho pretendido.

Definir a quantidade correta de poder velocidade computacional.

Existem análises que devem ser executadas em tempo real.

Uma quantidade de dados devem ser armazenados.

Definir a quantidade correta de redundância, para ficar protegido.



Introdução a Big Data

Definindo o Big Data

Estabelecendo a base de arquitetura

Quantos dados minha empresa precisa administrar agora e no futuro?

Com que frequência minha empresa precisará administrar dados em tempo real ou próximo do tempo real?

Qual a quantidade de risco que minha empresa consegue suportar?

O setor industrial da empresa está sujeito a exigências rígidas de segurança, confiabilidade e governança?

Quão importante é a velocidade para minha demanda de gestão de dados?

Quão certos ou precisos devem ser os dados?



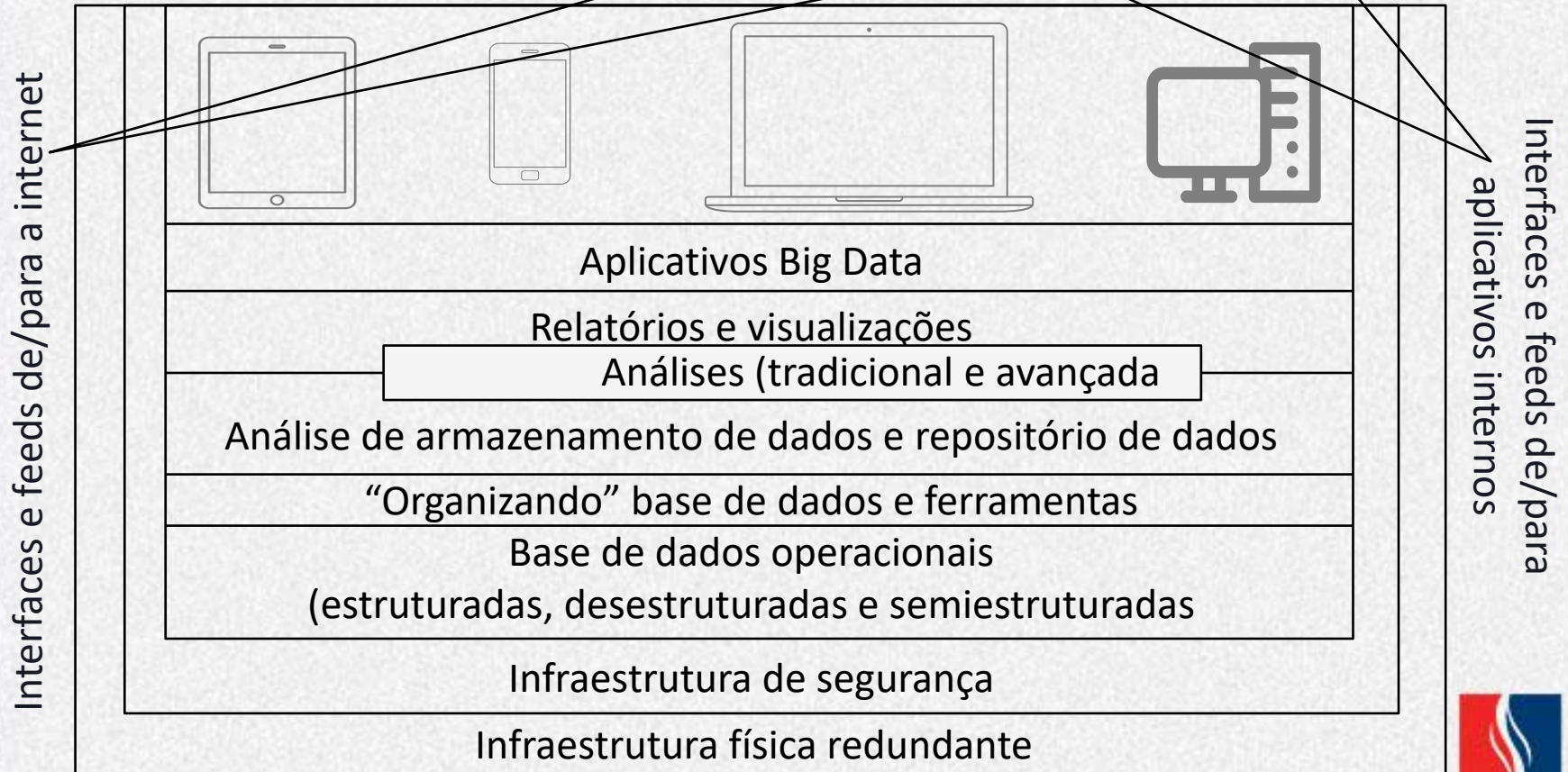
EDUCAÇÃO
METODISTA

Introdução a Big Data

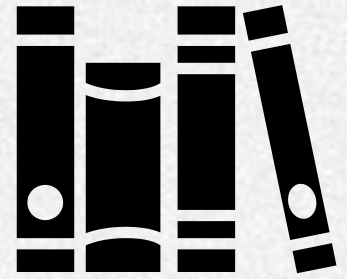
Definindo o Big Data

O que faz o Big Data Grande é o fato de ser baseado em coletar vários dados de diversas fontes. Interfaces de programação de aplicativos abertos serão o ponto central de qualquer arquitetura Big Data

Arquitetura Big Data



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