

# Gestão da Tecnologia da Informação

## Introdução a Big Data

Francisco José Tosi



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Aula 04 Conceitos Big Data

21/03

Francisco José Tosi



EDUCAÇÃO  
METODISTA





# Introdução a Big Data

## Definindo o Big Data

### Estabelecendo a base de arquitetura - Infraestrutura física redundante

A infraestrutura física de suporte é fundamental para a operação e para definir o tamanho de uma arquitetura de Big Data.



É necessário uma infraestrutura robusta para o Big Data suportar um volume de dados inesperado e imprevisível.

A infraestrutura física deve ser baseada em modelo de computação distribuída, ou seja, os dados podem ser armazenados fisicamente em locais diferentes e serem acessados através de rede.



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Definindo o Big Data

### Estabelecendo a base de arquitetura - Infraestrutura física redundante

A redundância é necessária por ser utilizado muitos dados de muitas fontes diferentes.



A redundância pode ser construída de várias formas:

- Criar uma nuvem particular, a redundância será criada no ambiente privado, de forma que seja possível suportar mudanças nas cargas de trabalho.
- Criar uma nuvem externa aumentando os recursos os recursos internos de TI.
- Utilizar um software como serviço (SaaS, Software as Service) permitindo a empresa realizar análise de dados como um serviço.



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## Definindo o Big Data

### Estabelecendo a base de arquitetura - Infraestrutura de segurança

A proteção de dados torna-se mais importante a medida que a análise do Big Data se importante para a empresa.



A proteção dos dados são necessários para atender as necessidades legais e também para proteger a integridade dos clientes.

Deve ser determinado quem pode acessar e quais circunstancias pode ser acessado.

Identificar quem são os clientes/usuários.

A segurança dos dados deve estar presente desde o inicio do projeto de criação do Big Data.



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Definindo o Big Data

### Estabelecendo a base de arquitetura - Infraestrutura de dados operacionais

Deve ser levado em conta todas as fontes de dados que fazem parte do negócio da empresa.

Identificar qual impacto das informações na forma que serão gerenciadas.

Atualmente os dados podem ser recebidos desestruturados e de várias fonte (ex. mídia social).

Em um ambiente Big Data pode existir arquitetura de banco dados documental, gráfica, colunar e geográficos.

Deve ser identificado a arquitetura dos dados e aos tipos de transações, garantindo que as informações serão acessadas quando for necessário.



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## Definindo o Big Data

### Estabelecendo a base de arquitetura - Infraestrutura de dados operacionais

Características comuns das fontes de dados:

- ✓ Elas representam sistemas de gravação que mantêm o controle dos dados essenciais necessários em operações diárias e em tempo real do negócio.
- ✓ Elas são atualizadas continuamente com base em transações acontecendo em unidades de negócio e a partir da web.
- ✓ Para essas fontes fornecerem uma representação precisa do negócio devem combinar dados estruturados e não estruturados.
- ✓ Esses sistemas também devem ser capazes de se adaptar para suportar milhares de usuários em uma base consistente, incluindo sistema de e-commerce transacional, sistemas de gestão de relacionamento com clientes ou aplicativos de *call center*



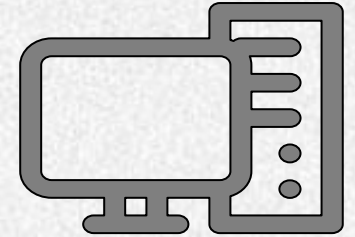
EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Definindo o Big Data

### Desempenho

A arquitetura de dados deve ser de acordo com a infraestrutura



*“Para determinar modelos para decisão onde será relacionado grandes volumes de informações diferentes, pode levar dias para executar se for utilizado um servidor tradicional. Porém pode levar minutos se for utilizado um modelo de computação distribuída”*

O tipo de base de dados pode ser determinada pelo desempenho necessário.

*“Para entender como dois elementos distintos estão relacionados, não poderia ser utilizado uma base de dados relacional, e sim uma base gráfica”*



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

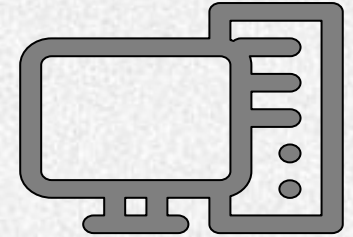
## Definindo o Big Data

### Desempenho

A utilização da base de dados correta melhora o desempenho

A base de dados gráfica é bem utilizada para aplicativos científicos e técnicos

A base de dados colunares é uma forma de melhorar a eficiência já que armazena informações em coluna ao invés de linhas



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Definindo o Big Data

## Desempenho





# Introdução a Big Data

## Definindo o Big Data

### MapReduce

Projetado pela Google



Seu objetivo é de executar eficientemente um conjunto de funções para processar uma quantidade grande de dados em modo lotes

**Map** distribui o problema de programação, ou tarefas, através de um grande numero de sistemas, equilibrando e administrando a recuperação de falhas



**Reduce** Agrega os elementos para oferecer o resultado

Ex. determinar quantas paginas de um livro são escritas em 50 linhas diferentes



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Definindo o Big Data

### Big Table

Desenvolvida pela Google



Sistema de armazenamento distribuído destinado a administrar dados estruturados altamente escaláveis.

É um mapa multidimensional persistente, esparsos e distribuído.

É desenvolvido para armazenar volumes enorme de dados através de servidores commodities.



As informações são organizadas em tabelas linhas e colunas, diferente do modelo de base de dados relacional tradicional.



# Introdução a Big Data

## Definindo o Big Data

### Hadoop



Plataforma de software gerenciado via Apache

Derivada do MapReduce e Big Table.

Permite que aplicativos com base em MapReduce executem grandes agrupamentos de hardware commodity.

Arquitetura de computação que apoia os negócios do Yahoo!.



Executa o processamento de dados em paralelo através de nós de computação para acelerar cálculos e latência oculta

# Introdução a Big Data

## Definindo o Big Data

### Hadoop

Existe dois componentes principais:

- ✓ Um sistema de arquivos distribuídos altamente escalável que pode suportar petabytes de dados .
- ✓ Um mecanismo de MapReduce altamente escalável que calcula resultado em lotes.



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## Definindo o Big Data

### Análise – Tradicional e Avançada

O que fazer com todos os dados apurados?

- ✓ Utilizar armazenamento de dados tradicional?
- ✓ Efetuar análise avançada?

**Administrar o Big Data abrangente requer abordagens diferentes a fim de conseguir melhor resultado para a empresa e ajudar o negócio planejar seu futuro**



**EDUCAÇÃO  
METODISTA**

# Introdução a Big Data

## Definindo o Big Data

### Armazém de dados e repositórios de dados analíticos

Após classificar as grandes quantidades de dados capturados, separar os subconjuntos de informações relevantes e coloca-los em formato que ajude a área de negócios a tomar decisões

Os armazéns de dados permitem compreensão, particionamento multinível e uma arquitetura massiva de processamento paralelo





# Introdução a Big Data

## Definindo o Big Data

## Análise de Big Data

Analisar grande quantidade de dados permite as empresas entender agrupamentos de informações que podem gerar vantagens competitivas

Mecanismos de análises que possam analisar os dados altamente distribuídos de forma eficaz e fornecer resultados otimizados para resolução de problemas

A análise dessa grande quantidade de dados pode ficar muito complexa

Análise que une dados estruturados e desestruturados para prever fraudes



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Definindo o Big Data

## Relatórios de Visualização

As empresas utilizam relatórios para entender os que os dados dizem

O Big Data muda a maneira de administrar e utilizar os dados

A análise agora não é realizada não apenas focando em elementos de dados, mas também como os elementos de dados criam contexto com base no problema a ser resolvido

Com o Big Data relatório e visualizações de dados se tornam ferramentas para análise de como os dados estão conectando e de que forma isso pode afetar o futuro



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## Definindo o Big Data

## Aplicativos Big Data

Com o Big Data, os aplicativos que são desenvolvidos são projetados para tirar vantagens sobre as características únicas do Big Data

Os aplicativos contam com enorme volume de dados, dados variados

Na **área de saúde** um aplicativo pode ser capaz de monitorar crianças prematuras e indicar se uma intervenção é necessária

Na **área de produção** um aplicativo pode prevenir o desligamento de uma máquina durante o ciclo de produção

Na **área de gestão de tráfego** um aplicativo pode reduzir o número de engarrafamento em rodovias

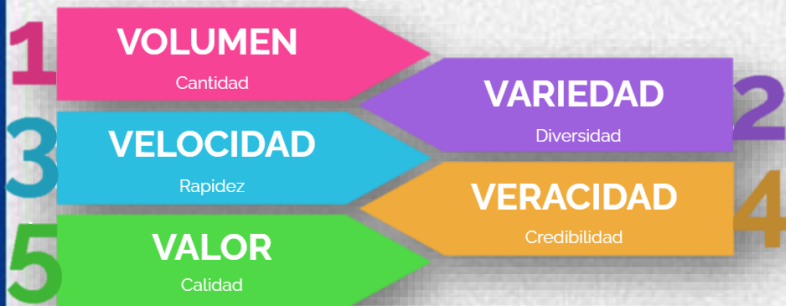
# Introdução a Big Data

big data são dados que contêm maior variedade, chegando em volumes crescentes e com mais velocidade



**big data** é um conjunto de dados maior e mais complexo, especialmente de novas fontes de dados.

São conjuntos de dados volumosos que o software tradicional de processamento de dados simplesmente não consegue gerenciá-los.



Esses grandes volumes de dados podem ser usados para resolver problemas de negócios que você não conseguiria resolver antes.



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## Os V's do Big Data



Volume

Com o big data, possível ter que processar grandes volumes de dados não estruturados de baixa densidade.

Podem ser dados de valor desconhecido, como feeds de dados do Twitter, fluxos de cliques em uma página da web ou em um aplicativo móvel, ou ainda um equipamento habilitado para sensores.

Para isso pode se necessário utilizar dezenas de terabytes de dados.



Ou até centenas de petabytes.



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Os V's do Big Data



Volume

Volume é uma variável subjetiva, pois os limites de armazenamento de dados aumentam com a rápida evolução tecnológica.

Corporações grandes, médias ou pequenas trabalham com dezenas de terabytes de dados .

O volume de dados utilizados pelas empresas tende a crescer geometricamente, sendo cada vez mais competitivo e rápido.

As empresas tem a necessidade de tomar decisões baseadas não apenas em palpites, mas sim em dados concretos



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## Os V's do Big Data



Velocidade

Velocidade é a taxa mais rápida na qual os dados são recebidos e talvez administrados.

Normalmente, a velocidade mais alta dos dados é transmitida diretamente para a memória, em vez de ser gravada no disco.

Alguns produtos inteligentes habilitados para internet operam em tempo real ou quase em tempo real e exigem avaliação e ação em tempo real.



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Os V's do Big Data



Velocidade

Atualmente é crucial saber tratar os dados na velocidade adequada.

Dados não tratados e analisados em tempo adequados são inúteis.

Estima-se que 90% dos dados digitais não estão sendo aproveitados adequadamente.

Dados passam a ser ativos corporativos importantes e deverão ser quantificados economicamente.

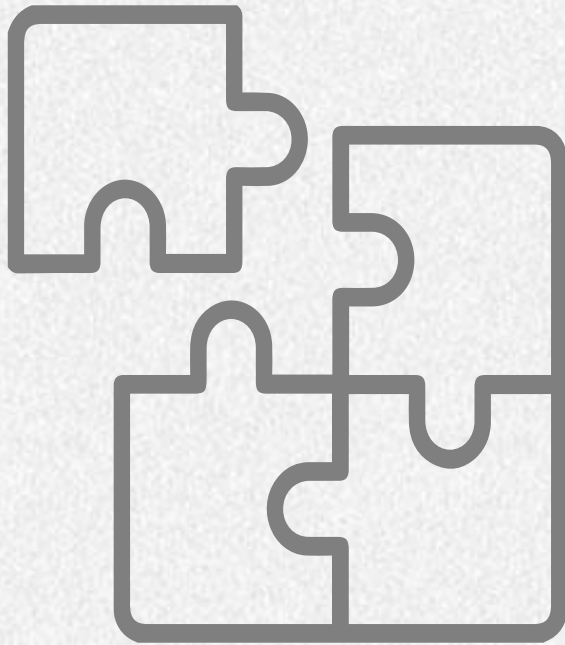


EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## Os V's do Big Data



Variedade

Variedade refere-se aos vários tipos de dados disponíveis.

Tipos de dados tradicionais foram estruturados e se adequam perfeitamente a um banco de dados relacional.

Com o aumento de big data, os dados vêm em novos tipos de dados não estruturados.

Tipos de dados não estruturados e semiestruturados, como texto, áudio e vídeo, exigem um pré-processamento adicional para obter significado e dar suporte a metadados.



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Os 5 V's do Big Data



### Veracidade

Para assegurar que a análise de dados seja capaz de atender às necessidades do negócio, é crucial que a empresa consiga trabalhar com conjuntos de dados confiáveis.

Muitas vezes os registros utilizados não são estruturados, o que pode levar a cenários em que o número de ruídos é alto e acaba impactando na qualidade do trabalho, as soluções de Big Data devem estar preparadas para buscar por dados de fontes confiáveis.

Deve existir a possibilidade de filtrar quais conteúdos são relevantes para o negócio, e eliminar os que não são confiáveis ou que não têm relevância.



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## Os V's do Big Data



Valor

Os dados devem ser capazes de agregar valor a processos e tornar os serviços mais competitivos.

Ao identificar tendências e padrões que gerem possibilidades para o gestor tomar decisões com confiança e direcionar melhor a estratégia para conquistar clientes e/ou mais mercados.

Pelo aspecto operacional, permitir a avaliação das rotinas internas e o uso de ferramentas corporativas, leva ao rastreamento de gargalos e torna a gestão de processos mais eficiente.



EDUCAÇÃO  
METODISTA

# Introdução a Big Data

## Os V's do Big Data

O Big Data só faz sentido se o valor da análise dos dados compensar o custo da sua coleta, armazenamento e processamento.



Valor



EDUCAÇÃO  
METODISTA



# Introdução a Big Data

## História do "Big Data"

O conceito de big data é relativamente novo, as origens de grandes conjuntos de dados remontam às décadas de 1960 e 1970, quando o mundo dos dados estava apenas começando, com os primeiros data centers e o desenvolvimento do banco de dados relacional.

Por volta de 2005, as pessoas começaram a perceber a quantidade de usuários de dados gerados pelo Facebook, YouTube e outros serviços on-line.

O Hadoop (uma estrutura de código aberto criada especificamente para armazenar e analisar grandes conjuntos de dados) foi desenvolvido no mesmo ano.

O NoSQL também começou a ganhar popularidade durante esse período.



# Introdução a Big Data

## História do "Big Data"

O desenvolvimento de estruturas de código aberto, como o Hadoop, (e, mais recentemente, o Spark) foi essencial para o crescimento do big data, porque elas tornaram o trabalho com big data mais fácil e seu armazenamento mais barato.

Nos anos seguintes, o volume de big data disparou. Usuários ainda estão gerando grandes quantidades de dados, mas não são somente humanos que estão fazendo isso.





# Introdução a Big Data

## Atividade 3

Realizar a leitura individual do artigo:

**O que é data lake e qual a sua importância?**

Disponível em:

[Data lake: o que é e qual a sua importância?  
\(indicium.tech\)](http://indicium.tech)

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

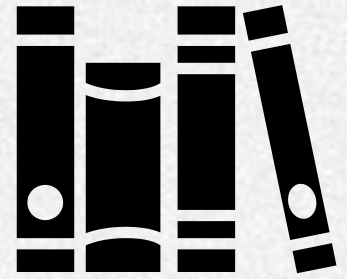
Em grupo produzir um resumo do artigo e disponibilizar no Moodle.

**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