

Gestão da Tecnologia da Informação

Introdução a Big Data

Francisco José Tosi



EDUCAÇÃO
METODISTA

Introdução a Big Data

Aula 08

Armazenamento e Análise de Dados

02/05

Francisco José Tosi



EDUCAÇÃO
METODISTA



Introdução a Big Data

Armazenamento e Análise de dados

Não podemos pensar em Big Data sem fontes de dados operacionais.

Uns dos serviços importantes para armazenamento de dados é a PERSISTÊNCIA

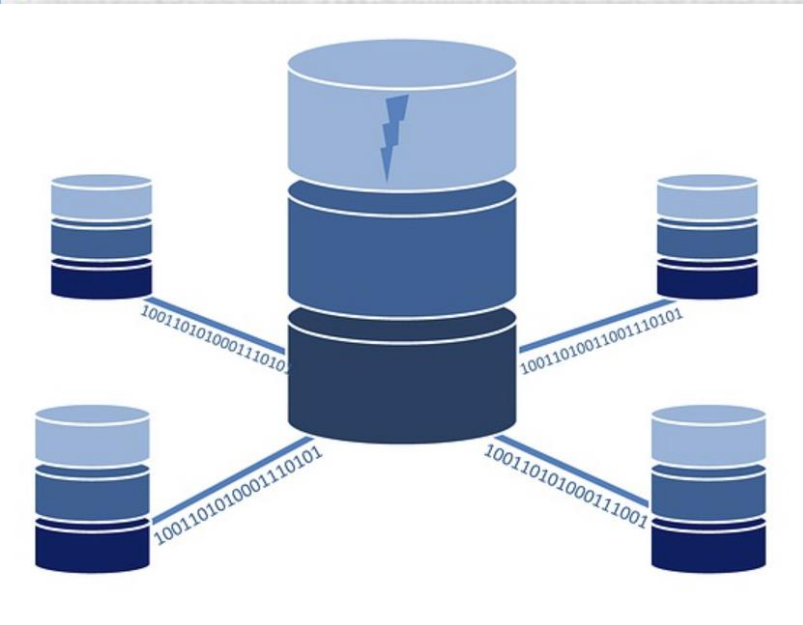
A **persistência** garante que os dados armazenados não sejam alterados por quem não tem autorização.

Garante também que sejam consultados quando necessário.

Uma base de dados só tem serventia se ela não for confiável para proteger os dados que armazena

Introdução a Big Data

Armazenamento e Análise de dados



É importante definir que tipo de dados deve ser persistido, como acessá-los e utilizá-los para tomada de decisão.

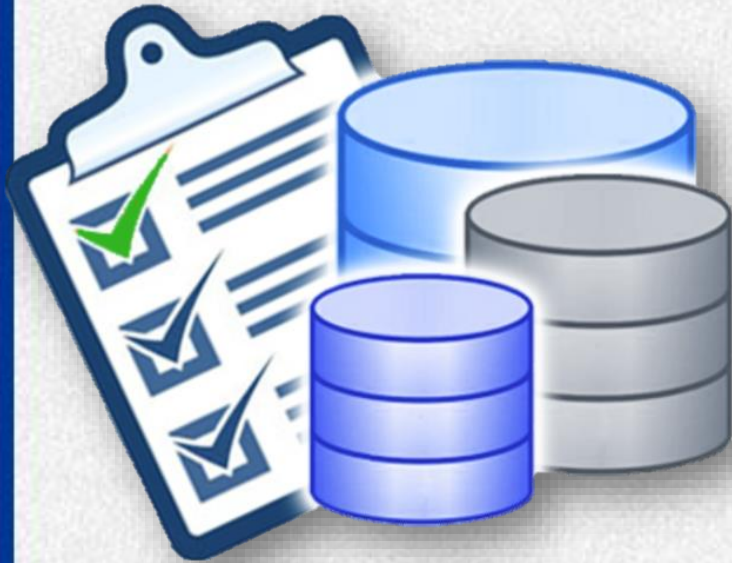
Devemos identificar também a escolha de mecanismos de base de dados.

Um armazenamento de dados persistentes é gestão de base de dados relacional RDBMC em português SGBDR (sistema Gerenciador de Banco de Dados Relacional).

Introdução a Big Data

Armazenamento e Análise de dados

Importância do RDBMS para o Big Data



A maioria dos dados das empresas são armazenadas em RDBMS.

São utilizados RDBMS diferentes para áreas diferentes com diferentes negócios.

É necessário depender de dados armazenados em RDBMS para criar níveis mais altos para o valor do negócio com Big Data

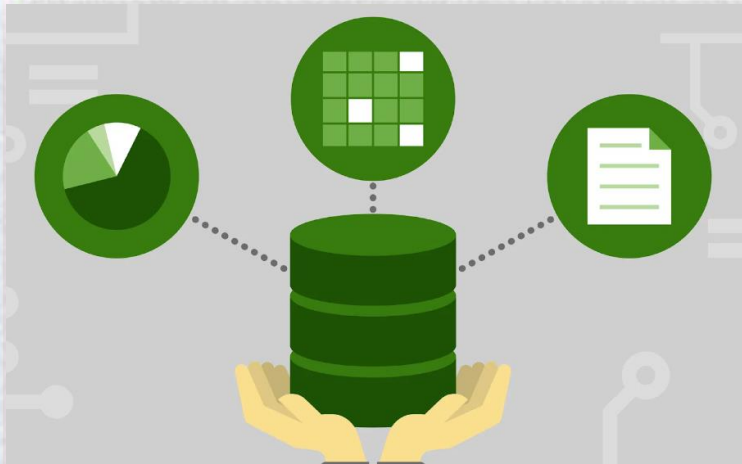


EDUCAÇÃO
METODISTA

Introdução a Big Data

Armazenamento e Análise de dados

Base de dados não relacionais



Banco de dados não relacional, não depende de tabela/chave.

Uma classe emergente é chamada de NoSQL – não apenas SQL.

São bases que não suportam o modelo relacional



EDUCAÇÃO
METODISTA

Introdução a Big Data

Armazenamento e Análise de dados

Base de dados não relacionais - características

Escalabilidade: capacidade de escrever dados simultaneamente através de múltiplos armazenamentos sem levar em consideração limitações físicas da infraestrutura da base. As bases de dados devem ser capazes de expandir e contrair conforme o fluxo de dados.

Modelo de dados e consulta: utilizam estrutura específica para armazenar informações com um conjunto exigido de APIs de consultas específicas para acessar dados de maneira inteligente.

Design e persistência: utilizam mecanismo diferentes para persistir os dados. A opção de mais alto desempenho é “em memória” onde a base inteira é mantida em um sistema de memória muito rápido de seus servidores



Introdução a Big Data

Armazenamento e Análise de dados

Base de dados não relacionais - características

Diversidade de interface: além das maiorias dos mecanismos das tecnologias elas oferecem uma ampla variedade de mecanismos de conexão para programadores e administradores de base de dados, incluindo ferramentas de análise e relatórios/visualização.

Consistência eventual: utiliza BASE (Basically Available Soft State e Eventual Consistency) estado suave e consistência eventual, onde consistency é responsável pela resolução e de conflitos quando os dados estão em movimento entre nós em uma implementação distribuídas. O estado dos dados é mantido pelo software e o modelo de acesso depende de disponibilidade básica



EDUCAÇÃO
METODISTA

Introdução a Big Data

Armazenamento e Análise de dados

Base de dados de Pares de Valor-Chave

Não exigem esquema como RDBMS e oferecem grande flexibilidade e estabilidade.

Não oferecem capacidade ACID (Atomicity, Consistency, Isolation, Durability – atomicidade, consistência, isolamento, durabilidade).

Exigem que os implementadores pensem sobre o posicionamento de dados, a replicação e tolerância a falhas, pois não são controlados pela tecnologia.

A maioria das informações é armazenada como *stings*.



Introdução a Big Data

Armazenamento e Análise de dados

Base de dados de Pares de Valor-Chave

Chave	Valor
Cor	Azul
Libação	Cerveja
Herói	Soldados
Chave	Valor
Usuário 1_cor	Vermelho
Usuário 2_cor	Marrom
Usuário 3_libação	Vinho
Usuário 4_Herói	Vendedor top do mes



Introdução a Big Data

Armazenamento e Análise de dados

Base de dados de Documentais

Existem dois tipos de bases documentais:

Documento completo Word, web pages, etc.

Base de dados para componentes de documentos

São úteis quando necessita de muitos relatórios que precisam ser montados dinamicamente a partir de documentos que mudam com frequência.

Ex. documentos da área de saúde em que a composição do muda com base no perfil da pessoa (idade, residência, renda) plano de saúde e elegibilidade para programa do governo.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Armazenamento e Análise de dados

Base de dados Colunares

São orientadas em linhas, porem os dados em cada linhas são armazenados juntos

Os dados são armazenados Através de linhas.

Existe a facilidade adicionar colunas, são adicionadas linha por linha, oferecem grande flexibilidade, desempenho e escalabilidade



EDUCAÇÃO
METODISTA

Introdução a Big Data

Armazenamento e Análise de dados

Base de dados Gráficas

Tem estrutura de relacionamento entre nós.

Útil quando tem que lidar com dados altamente conectados.

Pode ser utilizada para administrar dados geográficos.

Ex. exploração de petróleo, modelar e otimizar as redes de um provedor de telecomunicações



EDUCAÇÃO
METODISTA

Introdução a Big Data

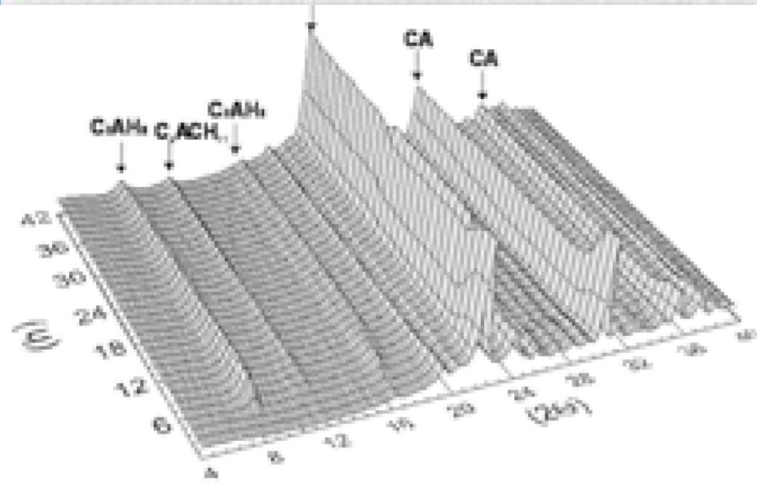
Armazenamento e Análise de dados

Base de dados Espaciais

São implementações do padrão OGC (Open Geofrafic Consortium).

São importantes para empresas que necessitam de se aproveitar de várias dimensões diferentes de informações para ajudar na tomada de decisão.

Pode ser utilizada para administrar dados geográficos.



Ex. um meteorologista fazendo uma pesquisa pode querer armazenar e avaliar dados relacionados a um furacão, como temperatura, velocidade do vento e umidade, e modelar os resultados em 3 dimensões.

Armazenam dados bidimensionais, tridimensionais.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Armazenamento e Análise de dados

O armazenamento de dados está relacionado com a utilização de mídias de gravação para retenção dos dados capturados.

Pode ser feito por meio de arquivos, blocos, e objetos.

Deve ser utilizado conforme a necessidade de utilização.

Garante a recuperação no futuro.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Armazenamento e Análise de dados

Para garantirmos um bom armazenamento devemos ter:

- ✓ Segurança.
- ✓ Integridade;
- ✓ Disponibilidade;
- ✓ Redundância reduzida;
- ✓ Controle de concorrência;
- ✓ Otimização de recursos; etc.



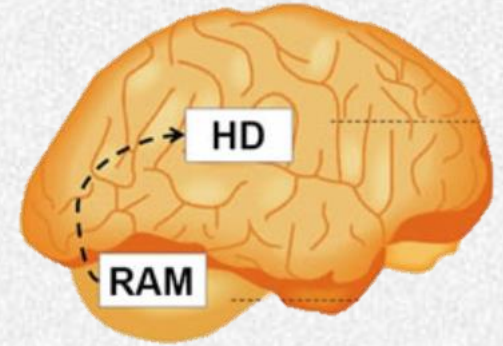
EDUCAÇÃO
METODISTA

Introdução a Big Data

Armazenamento e Análise de dados

Armazenamento em dispositivo volátil: mantém o estado do dado enquanto o dispositivo estiver ligado na energia, sem energia os dados serão perdidos.

Ex. Memória RAM



Armazenamento em dispositivo não volátil: tem a capacidade de manter os dados sem presença de energia.

Ex. Disco rígido



EDUCAÇÃO
METODISTA

Introdução a Big Data

Modelagem e correlação de dados

Big Data *Analytics*

É a capacidade de utilizar dados, realizar análise e utilizar um raciocínio sistêmico para conduzir um processo de tomada de decisão.

Utilizar a **Inteligência analítica** para melhorar o desempenho com relação aos domínios fundamentais do negócio por meio de dados e análise .

A *analytics* é um conjunto de tipos de análises sobre dados realizadas com a finalidade de obter indicadores de desempenho ou novas visões sobre os dados tratados.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Modelagem e correlação de dados

Análise descritiva e preditiva

Análise descritiva é para viabilizar decisões de negócio baseadas em fatos e dados e não em sentimentos pessoais.

Atualmente essa análise somente não é suficiente devido a sociedade gerar uma imensidão de informações tornando a tomada de decisão não precisa .

Análise preditiva foca no futuro define decisões de negócio e processos com mais amplitude mais objetiva.

Tira a empresa do universo particular e colocando diante de uma sociedade consumidora e global



EDUCAÇÃO
METODISTA

Introdução a Big Data

Modelagem e correlação de dados

Análise descritiva e preditiva

Análise preditiva não é nova, era um tema acadêmico a vários anos.

Atualmente tem relevância no segmento profissional de TI, onde cresceu com a grande quantidade de dados capturados pelas pessoas.

- ✓ Transações online
- ✓ Redes sociais
- ✓ Utilização de equipamentos e sensores
- ✓ Dispositivos móveis



EDUCAÇÃO
METODISTA

Introdução a Big Data

Modelagem e correlação de dados

IoT – Internet das coisas

IoT é a capacidade de capturar, analisar e transmitir dados para as coisas, aumentando sua utilidade.

A IoT esta mudando as decisões de gerenciamento das empresas.

Empresas de logística ou quem tem atividades ligadas a logística utilizam análises para e otimizar seu desempenho



Esta em teste nos Estados Unidos utilização de sensores e análise para detectar o som de tiros, possibilitando responder a incidentes com arma de fogo

Introdução a Big Data

Modelagem e correlação de dados

Análise de clique

A **análise de cliques** em um site na *web* permite a coleta, análise e geração de relatórios de dados agregados sobre as páginas visitadas.



Identifica também caminho realizado que é navegado – **Chamado de fluxo de cliques** (*ClickStream*).

Existem dois níveis de análise de fluxo de cliques



EDUCAÇÃO
METODISTA

Introdução a Big Data

Modelagem e correlação de dados

Análise de clique

Análise de tráfego: opera no nível de serviço e rastreia quantas páginas são acessadas por usuários, quanto tempo fica em cada página e a frequência de utilização do botão de retorno ou parada.

Quantidade de dados transmitidos antes que o usuário se mova.



Introdução a Big Data

Modelagem e correlação de dados

Análise de clique

Análise baseada em comércio eletrônico: utiliza dados do clique para determinar a eficácia do site para o mercado.

Foca nas páginas em que o comprador acessa, o que ele consulta, quais os detalhes, o que inclui ou tira da lista de compras.



Identifica quais itens compra, não se preocupando se o cliente pertence a algum programa de fidelidade, ou utilizar códigos de desconto, ou qual método de pagamento.



EDUCAÇÃO
METODISTA

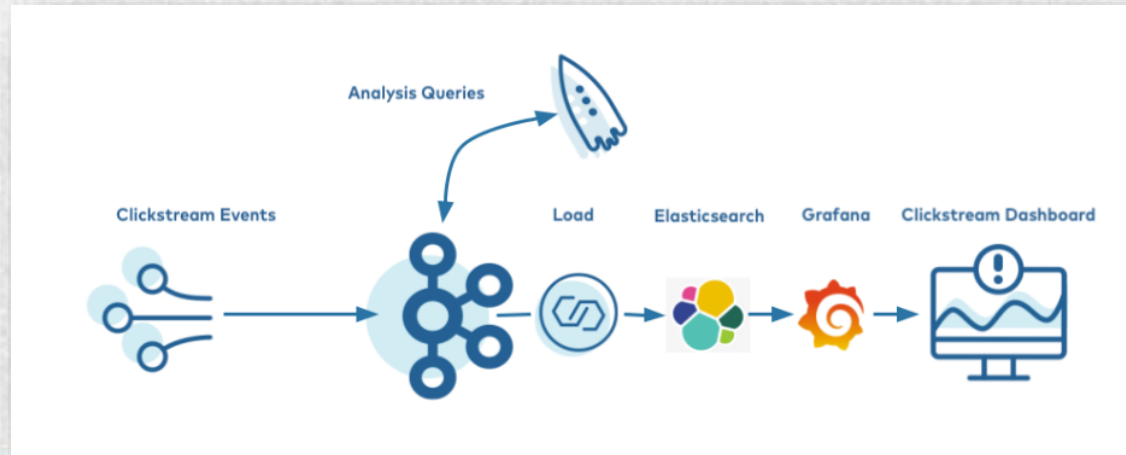
Introdução a Big Data

Modelagem e correlação de dados

Análise de clique

A **análise de fluxo de cliques** gera um grande numero de dados, portanto é necessário a utilização de ferramentas para ajuda.

Análise **Click Stream** é considerada mais eficaz quando usada em conjunto com outros recursos de avaliação.



Introdução a Big Data

Modelagem e correlação de dados

Análise de clique

A Amazon usa **Big Data** análise para detectar o que cada cliente adicionou ao seu carrinho de compras.

Faz relação do itens visitados ou comprados em um passado recente ou mais distante.

Técnica denominada *filtragem colaborativa item a item* foi criada por Greg Linden.

Utiliza fonte de dados estruturados e não estruturados para customizar a experiencia de compra dos usuários em um site.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Modelagem e correlação de dados

Modelagem

Análise preditiva é a tecnologia que faz uso da experiência (dados) para prever o comportamento dos indivíduos para melhorar as decisões

Para os sistemas industriais o valor da análise preditiva terá como o principal objetivo prever e prevenir problemas em produtos.



Introdução a Big Data

Modelagem e correlação de dados

Modelagem

Modelagem preditiva é uma área da estatística que trata da extração dos dados e de sua utilização para prever tendências e padrões de comportamento.

A análise preditiva pode ser aplicada a qualquer tipo de informação desconhecida, no passado, no presente e no futuro.

A análise preditiva se baseia na captura de relações entre várias expectativas as previstas em ocorrências passadas, para prever um resultado desconhecido.

Deve-se ter atenção, pois a precisão e usabilidade dos resultados dependerão do nível de análise e da qualidade dos dados.



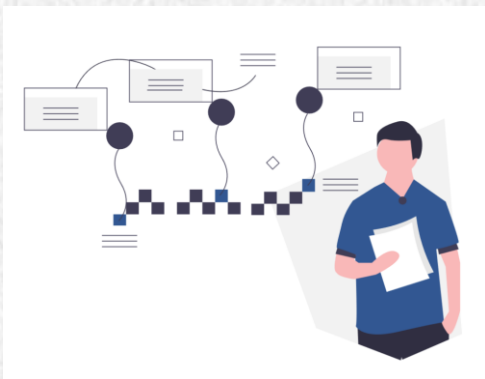
Introdução a Big Data

Modelagem e correlação de dados

Correlação de dados

A correlação de dados começou com o engenheiro de software Greg Linden, contratado pela Amazon para administrar o site da empresa.

A Amazon continha vários críticos e editores literários que selecionavam os livros e títulos apresentados na página e indicavam para os clientes, sem correlação, sem elementos que poderiam fazer o leitor se interessar pelo livro.



Greg percebeu que era melhor comparar os produtos e as associações entre eles.

Introdução a Big Data

Modelagem e correlação de dados

Correlação de dados



A **grande chave do Big Data** a descoberta da correlação entre os dados, que aparentemente nada tem a ver uns com os outros.

As correlações são fortes quando temos modificações do valor de alguns dados, o que faz com que o outro dado completamente diferente sofra alterações.

O princípio de correlação de dados associado a técnica *ClickStream* desencadeou a criação de algoritmos preditivos sobre a possibilidade de alguém vir se interessar por outro produto.

Essa correlação está bastante disseminada no *e-commerce* em compras iterativas



EDUCAÇÃO
METODISTA

Introdução a Big Data

Atividade 4

Realizar a leitura individual do artigo:

O que é Análise Preditiva e como utilizar na tomada de decisões?

<https://www.tecnicon.com.br/blog/481->

[O que e Analise Preditiva e como utilizar na tomada de decisoes](#)

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

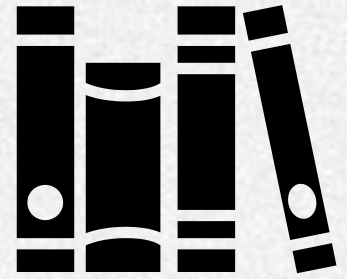
Em grupo produzir um resumo do artigo, fazer comentários sobre o que leu e disponibilizar no Moodle.

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