

Gestão da Tecnologia da Informação

Introdução a Big Data

Francisco José Tosi



EDUCAÇÃO
METODISTA

Introdução a Big Data

Aula 06

Fundamentos Tecnológicos para o Big Data

04/04

Francisco José Tosi



EDUCAÇÃO
METODISTA



Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Se tivermos que lidar com:

- Alto volume de dados transacionais;
- Requisitos de tolerância a falhas;
- Diferentes tipos de dados;
- Com fontes conhecidas e desconhecidas;

Construir um modelo extensível, conduzindo a um armazém de dados



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

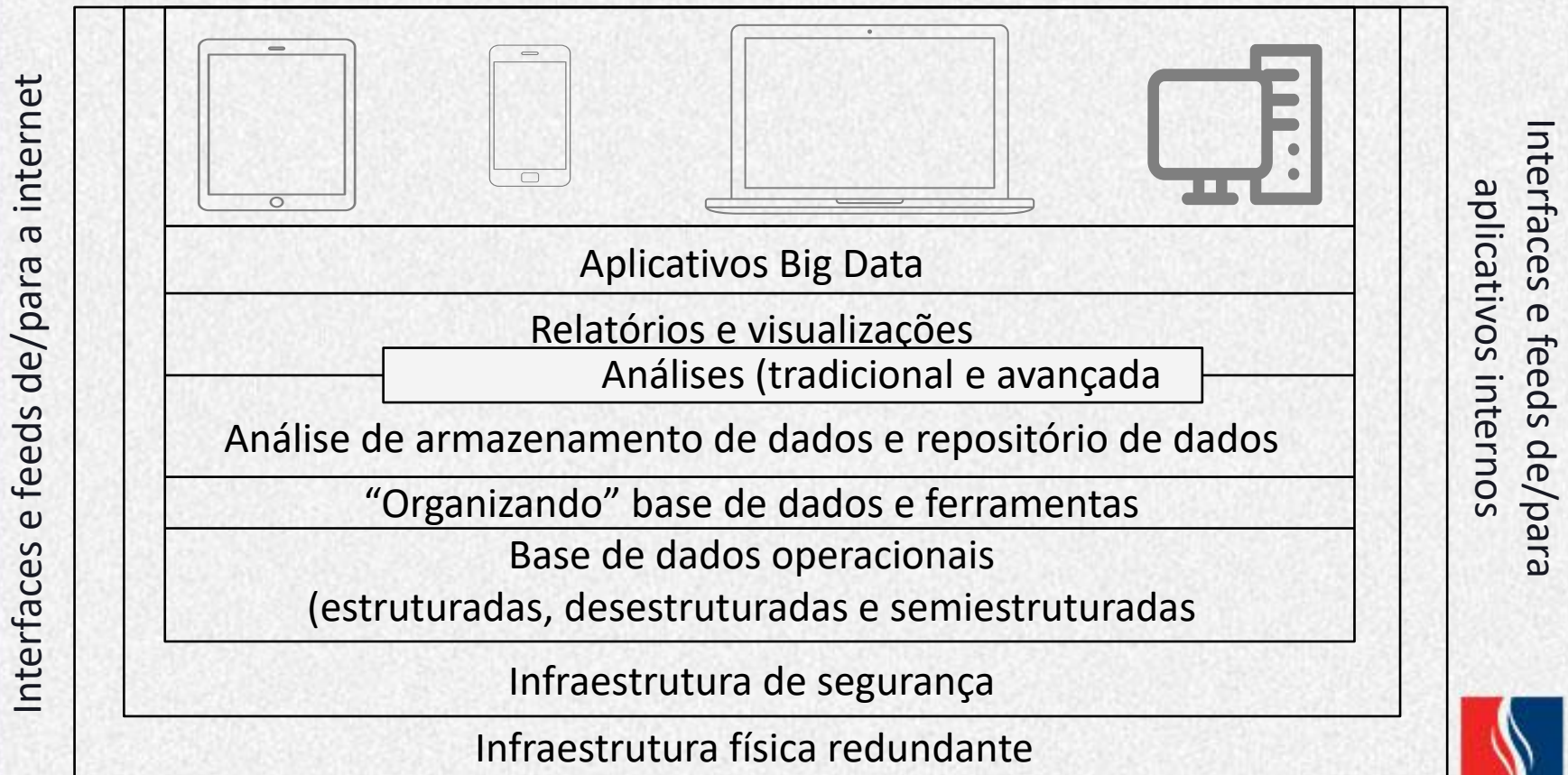
- A estrutura do Big Data deveria lidar com todas as exigências que discutimos:



Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data



Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 0 - Infraestrutura Física Redundante

Desempenho: Quanto receptivo o sistema precisa ser?

Latência deve ser medido de ponta a ponta, com base em uma única transação ou solicitação de consulta.

Alto Desempenho Baixa Latência tendem ser muito caros

Disponibilidade: É necessário a garantia de 100% do serviço?

Quanto tempo o negócio pode esperar no caso de interrupção ou falha?

Infraestrutura com alta disponibilidade é caro



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 0 - Infraestrutura Física Redundante

Escalabilidade: Qual o tamanho de infraestrutura é necessária?

Quanto espaço em disco é necessário agora e no futuro?

Quanto poder computacional é necessário?

Definir a necessidade e incluir um pouco mais para desafios inesperado

Flexibilidade: Com que rapidez consegue incluir mais recurso a infraestrutura?

Qual a velocidade o ambiente se recupera de falhas?

Infraestrutura flexível é cara, mas pode ser minimizada com serviços em nuvem



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 0 - Infraestrutura Física Redundante

Custo: Quanto de dinheiro está disponível?

Quanto espaço em disco é necessário agora e no futuro?

Uma estratégia é comprar a “melhor” rede e economizar no armazenamento, ou vice-versa

Estabelecer exigências para cada uma das áreas no orçamento global

Fazer compensações se for necessário



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 0 - Infraestrutura Física Redundante

O Big Data precisa ser de alta disponibilidade, para que as redes, os servidores, o armazenamento físico possam ser redundantes e resilientes.

Infraestrutura ou Sistemas, é resiliente a fracassos ou mudanças quando há recursos redundantes suficientes, estando posicionados e prontos para entrar em ação.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 0 - Infraestrutura Física Redundante

Exemplo:

Se existir apenas uma conexão de rede entre o negócio e a internet, não existe redundância a uma interrupção da rede.

Em grandes Data Centers com requisitos de continuidade de negócios, a redundância acontece e pode ser utilizada para criar o ambiente do Big Data.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 0 - Infraestrutura Física Redundante

As redes devem ser redundantes

Capacidade suficiente para acomodar o volume necessário

Velocidade tráfego normal



Em um Big Data é normal a
velocidade e o volume aumente



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 0 - Infraestrutura Física Redundante

Gestão de hardware: Armazenamento e servidores

Devem ter velocidade suficiente

Ser suficiente para lidar com a capacidade esperada do
Big Data



Não adianta ter rede de alta velocidade se o servidor é lento
Do mesmo modo não adianta ter servidores rápidos que ultrapassam o desempenho da rede



EDUCAÇÃO
METODISTA

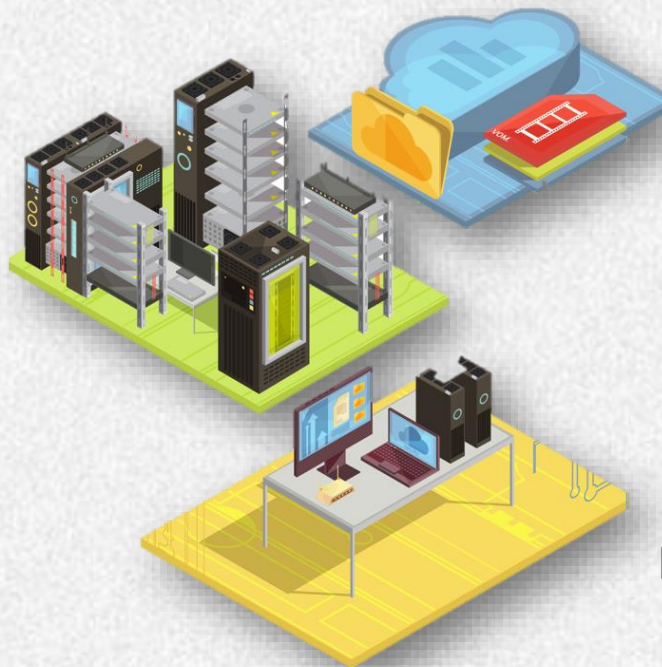
Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 0 - Infraestrutura Física Redundante

Operações de infraestrutura



Ambiente deve ser bem gerido para conseguir maiores níveis de desempenho

Ser capaz de prever e prevenir falhas

Deve ser seguro com integridade dos dados e os processos de negócios



EDUCAÇÃO
METODISTA

Introdução a Big Data

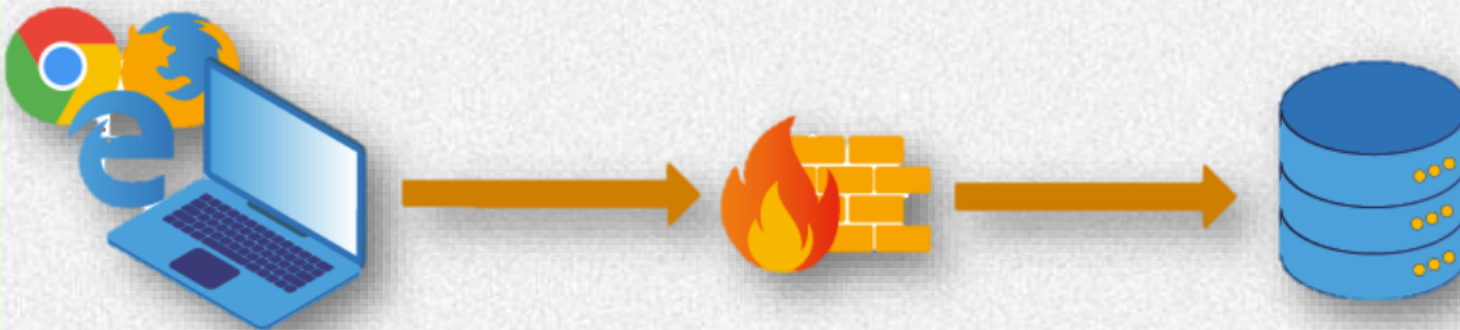
Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 1 - Infraestrutura de Segurança

Acesso a dados: Os dados devem estar disponíveis apenas para quem tem o direito e necessidade legítima de negócio

Acesso a aplicativos: manter controle sobre uso não autorizado



Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 1 - Infraestrutura de Segurança

Criptografia de dados: Criptografar grandes volumes de dados é muito caro e estressante, uma abordagem mais moderada é identificar elementos de dados que exigem um nível maior de segurança



Deteção de ameaças: Garantir vários perímetros de segurança, contra ameaças em dispositivos móveis e redes sociais

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 1 - Infraestrutura de Segurança

Interfaces e Fontes para e de Aplicativos e Internet



Devido ao grande volume de dados ser desestruturados o Big Data exige uma diferença em desenvolvimento de API

Uma nova técnica “LINGUAGEM NATURAL”
PLN

Fazer o interface entre o Big Data e os aplicativos



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 2 – Base de dados Operacionais

No centro do Big Data estão os mecanismos de base de dados com as informações relevantes ao negócio.



Mecanismos necessitam ser rápidos; escaláveis e extremamente sólidos.

Existem muitas tecnologias diferentes de base de dados, é muito importante que seja escolhida a que mais atenda as necessidades.

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 2 – Base de dados Operacionais

Características de base de dados:

Atomicidade: Uma transação é “tudo o nada” quando é atômica. Se qualquer parte da transação ou sistema de base falha, tudo falha.

Consistência: Apenas transações com dados válidos serão executadas na base de dados. Se os dados estão corrompidos ou são impróprio, ela não será completada e os dados não serão escritos na base.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 2 – Base de dados Operacionais

Características de base de dados:

Isolamento: Transações múltiplas e simultâneas não vão interferir umas nas outras. Todas as que forem válidas serão executadas até serem concluídas e na ordem que foram submetidas para processamento.

Durabilidade: Depois que os dados da transação são escritos na base de dados, eles ficam lá para sempre.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 2 – Base de dados Operacionais

Características de base de dados:

Mecanismo	Linguagem de consulta	Map Reduce	Tipo de Dados	Transações	Exemplos
Relacional	SQL, Python, C	Não	Typed	ACID	PostgreSQL; Oracle; DB/2
Colunar	Ruby	Haddop	Predefinido; Typed	Sim se habilitada	HBase
Gráfico	Walking; Search; Cypher	Não	Untyped	ACID	Neo4J
Documental	Commands	JavaScript	Typed	Não	Não MongoDB; CouchDB
Valor-chave	Lucene; Commands	JavaScript	BLOB; semityped	Não	Não Riak; Redis



Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 3 – Serviços de Organização de Dados e Ferramentas



Capturam, validam e reúnem vários elementos
Big Data

MapReduce é uma técnica muito utilizada

Muitos serviços de organização de dados são
mecanizados

São serviços utilizados para **coletar e organizar os** dados para
processamentos futuros

Precisam fornecer **integração, tradução, normatização e escala**

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Camada 3 – Serviços de Organização de Dados e Ferramentas

Sistema de arquivos distribuídos: acomodar a decomposição de fluxos de dados para fornecer escala e capacidade de armazenamento

Serviço de serialização: armazenamento de dados persistentes e chamadas de procedimentos remotos em várias linguagens

Serviço de coordenação: construção de aplicativos distribuídos

Ferramenta de extração, transformação e carga (ETL): carregamento e conversão de dados estruturados e desestruturados para o Hadoop

Serviço de fluxo de trabalho: agendamento de processos e fornecer estrutura de sincronização dos elementos do processo

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

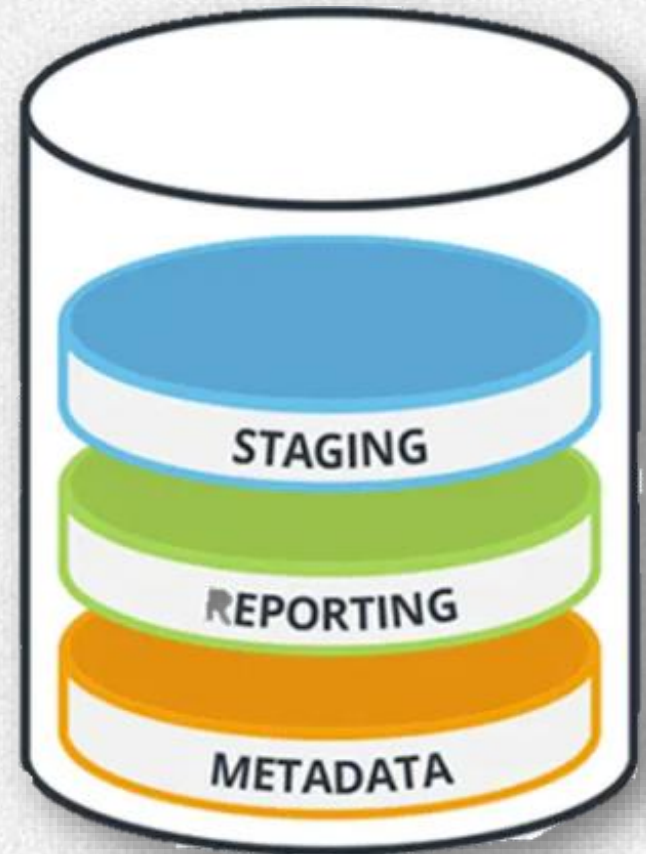
Explorando a pilha do Big Data

Camada 4 – Armazéns de Dados Analíticos

Armazéns de dados e repositório de dados tem sido a técnica utilizada para otimizar dados e ajudar a tomar decisões

Contém dados normalizados, coletados de uma variedade de fonte

São criados a partir de base de dados relacionais, dados multidimensionais, arquivos simples e base de dados de objetos tradicionais



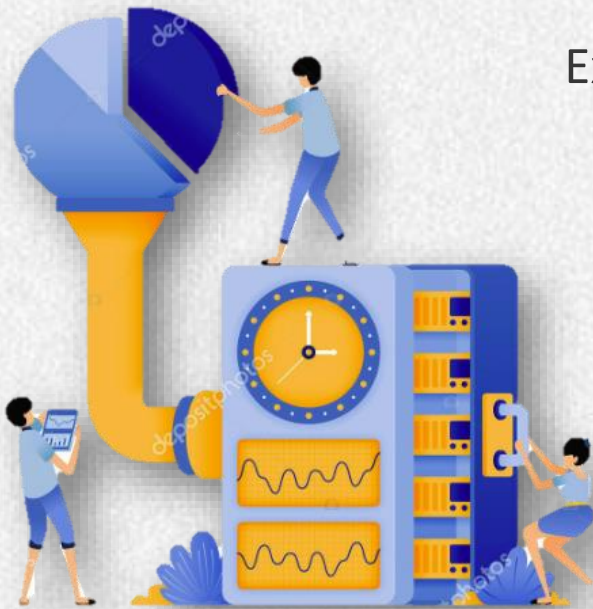
EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Análise Big Data



Existem ferramentas que ajudam na análise

Algoritmos devem ser capazes de trabalhar grande volume de dados em tempo real e com dados inconsistentes



EDUCAÇÃO
METODISTA

Introdução a Big Data

Fundamentos tecnológicos do "Big Data"

Explorando a pilha do Big Data

Análise Big Data

Classes de ferramenta para ajudar na análise Big Data

Relatórios e Painéis: Disponibilizam uma representação “de fácil utilização” da informação em várias fontes.

Visualização: Saída tende a ser altamente iterativa e dinâmica. Pode conter uma variedade de animação. Ex. mapas mentais, mapas de calor, infográficos e diagramas de conexão.

Análise e análise avançada: Atingem o armazém de dados e os processam para consumo humano. Devem explicar a tendências ou eventos que são transformativos, únicos ou revolucionários para a prática do negócio.



EDUCAÇÃO
METODISTA

Introdução a Big Data

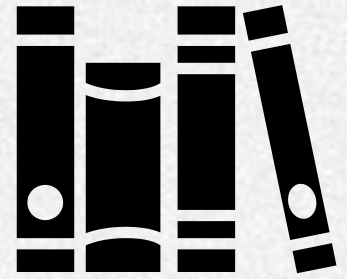
Fundamentos tecnológicos do "Big Data"

Boa 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