

Gestão da Tecnologia da Informação

Introdução a Big Data

Francisco José Tosi



EDUCAÇÃO
METODISTA

Introdução a Big Data

Aula 11

Equipe e Ferramentas para o Big Data

23/05

Francisco José Tosi



EDUCAÇÃO
METODISTA

Introdução a Big Data

Ferramentas Big Data

Ingestão de dados

O processo de consumir os dados necessários de forma adequada, eficiente e metódica.

Carga de lotes É possível acessar todos os tipos de dados necessários e dimensionar com eficiência o desempenho da carga de lotes nos repositórios de dados?

Captura de dados de mudanças É possível capturar as alterações feitas nos dados dos sistemas de origem sem afetá-los?

Fluxo de dados É possível coletar dados de forma confiável em tempo real e os transmite nos repositórios de dados?

Arquivamento É possível arquivar e compacta dados que não são usados com frequência, garantindo fácil acesso aos dados arquivados, se necessário?

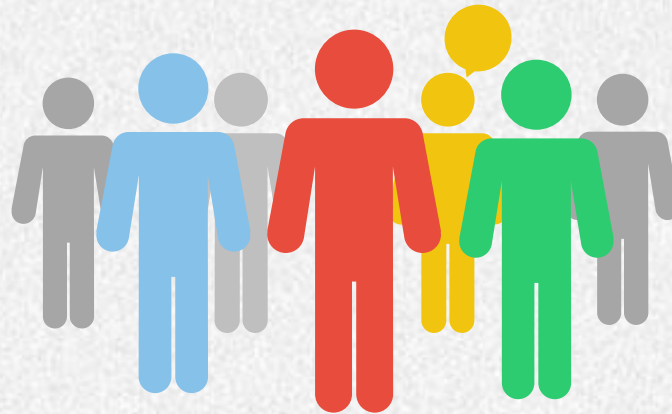


Introdução a Big Data

Criando equipe para o Big Data

A equipe de Big Data representa o maior desafio e a maior oportunidade.

É necessário um bom equilíbrio entre pessoas que entendem as metas de negócios e pessoas que podem executar suas exigências técnicas.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Criando equipe para o Big Data

Use as habilidades dos profissionais que você contratou



Um dos maiores erros que as empresas cometem quando contratam especialistas de dados e analistas quantitativos é pedir que façam o trabalho pesado.

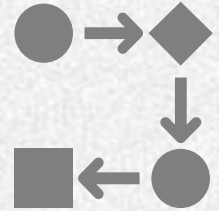
Não utilizem recursos mais qualificados o tempo todo em integrações de dados que exigem codificação manual e limpeza de dados.

Concentre as habilidades raras em tarefas que realmente precisam delas.

Não corra o risco de seus melhores profissionais pedirem demissão

Não permita que os melhores recursos percam tempo com um trabalho que pode muito bem ser feito com ferramentas.

Introdução a Big Data



Criando equipe para o Big Data

Crie a equipe estrategicamente

Quando o projeto começa dar certo, começa a crescer em escopo e recursos.

Deve ser pensado de maneira estratégica para ter consciência que não pode executar determinados processos com a rapidez necessária porque há um número limitado de pessoas habilitadas.

Se o projeto crescer em escopo, quais são as competências que podem ser encontrar a tempo de atender às suas necessidades.

Por exemplo, especialistas de dados são infinitamente mais difíceis de encontrar, treinar e contratar do que desenvolvedores.

Procure encontrar combinação certa de experiência em gerenciamento de dados conquistada com trabalho duro e do entusiasmo para aprender novas ferramentas.



**EDUCAÇÃO
METODISTA**

Introdução a Big Data

Criando equipe para o Big Data

Alinhar metas do projeto no início e depois as comunique



Um dos erros mais comuns que as empresas cometem ao contratar uma nova equipe é esquecer de comunicar as verdadeiras metas do projeto.

É preciso ficar claro o que está tentando oferecer aos usuários de negócios.

Potencialize o apoio dos executivos para divulgar a missão e compartilhar histórias de sucesso, bem como problemas.

Sem um controle firme do valor comercial do projeto, os novos contratados correm o risco de achar que só precisam pensar nas metas de TI do projeto



**EDUCAÇÃO
METODISTA**

Introdução a Big Data

Criando equipe para o Big Data

Quando a equipe aumenta, a necessidade de gerenciá-la também aumenta



O novo pessoal precisa se acostumar ao local de trabalho, ao trabalho que estão fazendo e às razões por que estão fazendo esse trabalho

A nova equipe precisa ser gerenciada de forma correta.

Elementos como cultura e coerência não podem ser subestimados.

Identifique como integrar as novas contratações em seus processos.

Talvez seja possível treiná-los para as habilidades, mas com certeza pode ajudá-los a serem melhores integrantes da equipe



**EDUCAÇÃO
METODISTA**

Introdução a Big Data

Criando equipe para o Big Data

A equipe não pode ficar parada



O big Data representa um desafio definitivo de começar antes dos concorrentes

As habilidades da equipe precisam evoluir com a mesma velocidade das mudanças que estão ocorrendo no mundo.

O que mais motiva os melhores profissionais é o desafio de chegar primeiro.

O desafio está em oferecer o treinamento e o debate necessários para continuar aumentando as habilidades dessas pessoas e as suas próprias.



**EDUCAÇÃO
METODISTA**

Introdução a Big Data

Ferramentas Big Data

Gerenciamento de dados

Todas as políticas, processos e práticas necessárias para gerenciar adequadamente a eficácia, a precisão, a confiabilidade e a disponibilidade dos dados.

Integração de dados É possível conseguir preparar e consolidar várias estruturas e fontes em um conjunto de dados consistente para análise?

Qualidade de dados É possível conseguir, de forma confiável, limpar os dados, identificar e remover os erros?

Segurança de dados É possível descobrir e proteger os dados em todos os repositórios de dados, atribuindo para isso regras sobre uso, acesso e permissões?

Máquina virtual de dados É possível conseguir criar uma camada de abstração para os dados que combine vagamente o processamento de dados no ambiente de implantação subjacente?



Introdução a Big Data

Ferramentas Big Data

Gerenciamento de dados

Todas as políticas, processos e práticas necessárias para gerenciar adequadamente a eficácia, a precisão, a confiabilidade e a disponibilidade dos dados.

Gerenciamento de dados mestres É possível armazenar uma versão consolidada, completa e fidedigna da verdade para vários domínios de dados?

Estrutura distribuída de dados É possível usar tecnologias como o Hadoop para dimensionar de forma econômica suas necessidades de armazenamento e processamento?

Data Warehouse É possível ter uma tecnologia de data warehouse que possa lidar com os requisitos de desempenho, uso e dimensionamento para análises e integrações de Big Data com infraestruturas Hadoop?



Introdução a Big Data

Ferramentas Big Data

Análise

As ferramentas e os processos que transformam dados brutos em insights, padrões, previsões e cálculos sobre o domínio que você está analisando.

Visualização É possível apresentar dados e resultados de maneira a facilitar a compreensão e o entendimento?

Análise avançada É possível aplicar algoritmos analíticos avançados aos seus conjuntos de dados para fazer cálculos complexos?

Aprendizado de máquina É possível conseguir aplicar sofisticados algoritmos de aprendizado de máquina para identificar padrões e fazer previsões em um nível para o qual não tem a largura de banda manual compatível?



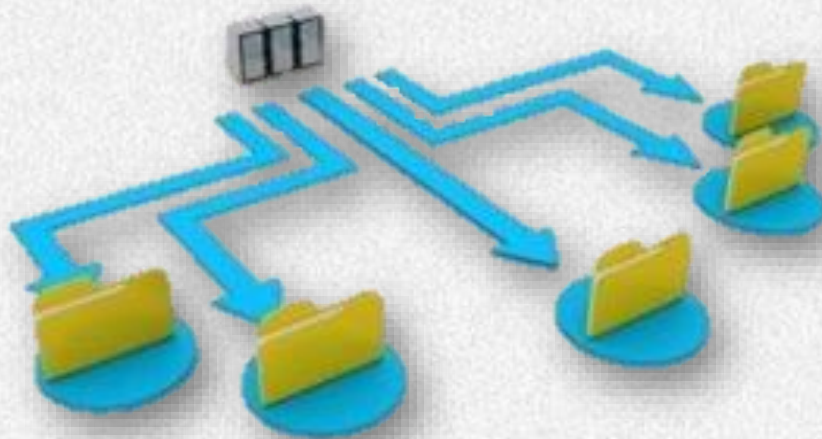
Introdução a Big Data

Oito processos para Big Data

1 – Acessar dados

Seu primeiro desafio será adquirir todos os dados de que precisa.

Em alguns casos, isso implicará a captura de dados de streaming e, em outros, significará extrair dados de um banco de dados.



Configure processos gerenciáveis e que possam ser repetidos para garantir o armazenamento desses processos de acordo com o uso que fará deles.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Oito processos para Big Data

2 – Integrar dados

O desafio mais complexo do Big Data envolve a variedade de estruturas e formatos de dados.



Para que a sua análise seja conduzida de forma sustentável, você precisará configurar um processo de integração e normalização de todos esses dados.

Idealmente, isso deve usar um processamento que seja o menos manual possível.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Oito processos para Big Data

3 – Limpar dados

Para que a sua análise seja confiável, é preciso garantir a limpeza dos dados a fim de eliminar duplicações, erros, imprecisões e dados incompletos.



O processo deve garantir que analistas e especialistas mais qualificados não fiquem o tempo todo fazendo o trabalho errado.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Oito processos para Big Data

4 – Criar dados mestres

Uma forma de manter uma fonte confiável de dados limpos e integrados é estabelecer um processo de criação de dados mestres.



O objetivo é criar uma coleção avançada de dados consolidados, organizada por domínio (como produtos, clientes etc.) e repleta de insights de Big Data, e que, assim, possa alimentar todos os outros sistemas.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Oito processos para Big Data

5 – Proteger dados

Aqui, você estabelecerá dois processos básicos. O primeiro será sobre como definir as regras e as práticas de segurança que cada conjunto de dados exige.



O segundo será sobre como detectar dados confidenciais e mascará-los de forma persistente ou dinâmica para garantir que essas regras e práticas recomendadas sejam aplicadas de maneira consistente.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Oito processos para Big Data

6 – Analisar dados

O processo de análise dependerá dos analistas, das ferramentas de análise e das necessidades relacionadas às metas.



A mentalidade de descoberta iterativa e melhoria contínua desempenhará um papel crucial nesse caso, pois esse processo deverá ficar melhor, mais rápido, mais barato e mais dimensionável ao longo do tempo e com uma experiência maior.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Oito processos para Big Data

7 – Analisar as necessidades de negócios

Esta etapa é crítica, mas quase sempre esquecida.



Configure um processo claro para a análise das necessidades dos negócios, mesmo quando estiver analisando os dados.

Isso é muito importante porque, caso você desacelere o ritmos dos negócios, poderá isolar seus esforços e minimizar o impacto nos próprios negócios.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Oito processos para Big Data

8 – Operacionalizar insight

O impacto do seu projeto de Big Data nos negócios precisa ser sentido.

Crie pipelines automatizados para as respostas que você encontrar e os entregue aos usuários de negócios que mais precisam deles.



Por exemplo, os dados sobre os clientes com maior probabilidade de rotatividade devem ser disponibilizados aos agentes de atendimento ao cliente por meio de um painel de controle. Lembre-se de incorporar um ciclo de feedback também, assim poderá ver como o insight é recebido.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dicas importante



Para que a cadeia de suprimentos de Big Data seja enxuta e eficaz, é preciso garantir que a arquitetura seja sólida e construída de maneira estratégica.

O ponto de partida mais Lógico para o Big Data, é a criação de um ambiente de desenvolvimento em uma área restrita, na qual você possa usar dados de teste para garantir a viabilidade da arquitetura.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dicas importante



Comece pequeno: Começando com uma área restrita bem definida sobre a qual tenha total controle, será possível conseguir a interação do caminho para uma implementação mais bem-sucedida. Documente as lições aprendidas em cada iteração.

Tamanho é importante: A diferença entre a área restrita e a implementação real é que o ambiente de produção será muito maior, exigindo um tratamento automatizado para ingerir, integrar, limpar e distribuir a saída. Será preciso ter uma estrutura muito mais robusta, além de componentes e processos comprovada e verdadeiramente confiáveis e flexíveis em um ambiente de produção dinâmico.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dicas importante



Faça o mascaramento antes de testar: Quando as organizações usam dados de teste, elas geralmente usam uma variante dos dados de produção dinâmicos para garantir que os formatos e as estruturas representem o ambiente dinâmico, se esses dados não forem mascarados de forma adequada, os dados confidenciais poderão ficar expostos em um ambiente de teste totalmente inseguro.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Dicas importante



Não se perca na tradução: Uma das fontes mais comuns de atrasos e desvios orçamentários onerosos dos projetos de Big Data decorre do fato de que os erros de codificação manual aparecem quando a arquitetura é ativada.

Quando for codificado manualmente partes significativas da arquitetura, espere o refatoramento de vários códigos para atender aos requisitos no nível de produção e administre as expectativas adequadamente.

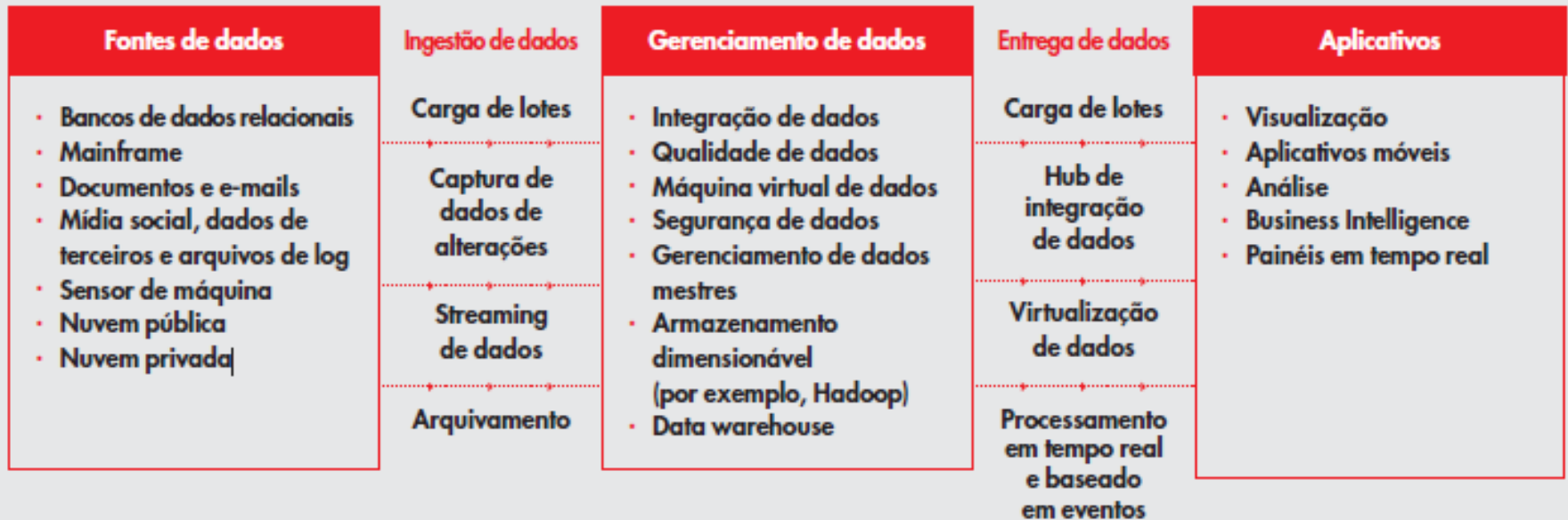
Como alternativa, pode ser utilizado ferramentas de produtividade e automação para evitar a necessidade de refatorar o código e os erros.



EDUCAÇÃO
METODISTA

Introdução a Big Data

Uma arquitetura para o Big Data



Introdução a Big Data

Atividade 9

Realizar a leitura individual do artigo:

As 10 melhores ferramentas de Big Data de código aberto para 2020

Disponível em:

<https://tecnetit.com.br/as-10-melhores-ferramentas-de-big-data-de-codigo-aberto-para-2020/>

NÃO será necessário a entrega da atividade, porém o assunto será abordado na prova.

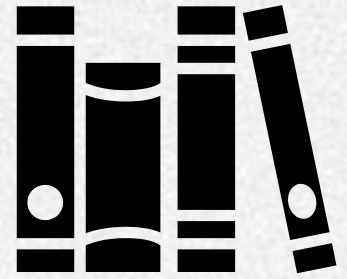


O Assunto abordado neste artigo será pedido em prova



**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