

Gestão da Informação e Ciclo de Vida Big Data

Karina Moura <kms2@cin.ufpe.br>

Roteiro

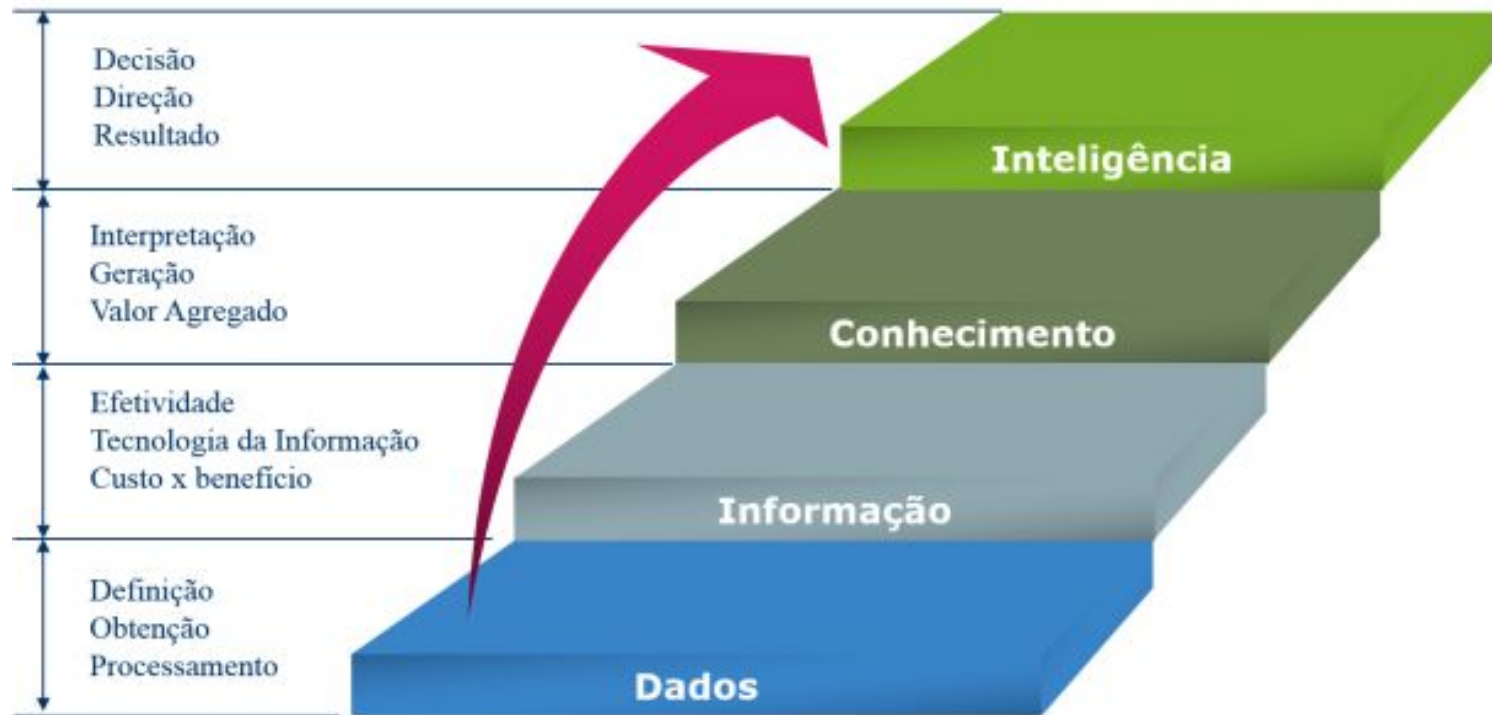
- Gestão da Informação
- Gerenciamento do ciclo de vida da informação
 - Governança
 - Tecnologias
 - Políticas de Gerenciamento
- Gerenciamento do ciclo de vida da informação para Big Data
 - Dados de Mídias Sociais
 - Política de gestão do ciclo de vida da informação de Mídias Sociais
- Benefícios da gestão do ciclo de vida da informação
- Análise de Big Data e Visualização de Dados
 - Análise de Big Data
 - Visualização de Dados

Gestão da Informação?

— — —

- Busca
- Identificação
- Classificação
- Processamento
- Armazenamento
- Disseminação de informações,

Fazer com que as informações cheguem às pessoas que necessitam delas para tomar decisões no momento certo



Gerenciamento do ciclo de vida da informação

“Prática de analisar o ciclo de vida dos dados numa empresa desde a sua criação ou aquisição até ao arquivamento.”

Porquê gerir dados?

Gerar Valor

Gerenciamento do ciclo de vida da informação

— — —

Gestão do ciclo de vida da informação constitui **um dos pilares fundamentais** na gestão dos dados numa empresa

É a plataforma na qual os três pilares da gestão de dados são projetados.

The diagram consists of a light gray horizontal bar containing three colored rectangular boxes. From left to right: a cyan box with the word 'Processo', a magenta box with the word 'Pessoas', and a yellow box with the word 'Tecnologia'. A thin black line with a downward-pointing hook is positioned above the center of the gray bar, and a larger black bracket is positioned above the entire diagram area.

Processo

Pessoas

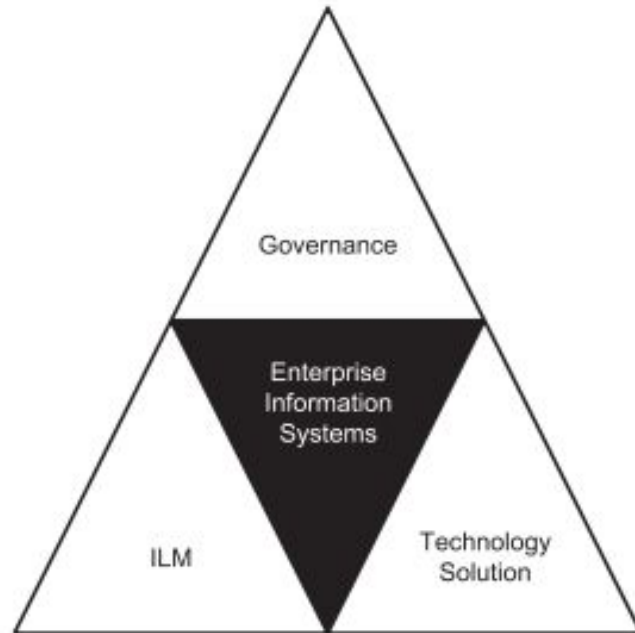
Tecnologia

Gerenciamento do ciclo de vida da informação

Objetivos:

- Gestão de dados como uma função empresarial.
- Melhorar as eficiências operacionais de sistemas e processos.
- Reduzir o custo total de propriedade, simplificando o uso de hardware e recursos.
- Ganhos de produtividade, reduzindo erros e melhorando a produtividade geral, automatizando o gerenciamento de dados e o ciclo de vida.
- Implementar um sistema auditável.
- Reduzir o risco de falha do sistema.
- Fornecer continuidade de negócios.
- Manter a flexibilidade para adicionar novos requisitos.

Gerenciamento do ciclo de vida da informação



Governança

— — —

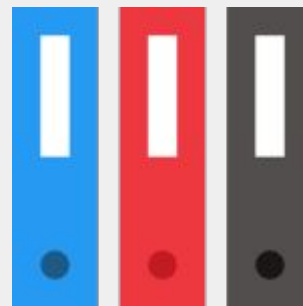
Governança de Informação

Lida com a criação de modelos de governança para os dados dentro da empresa



Governança de Programas

Lida com a implementação das políticas e processos estabelecidos na governança da informação



Governança

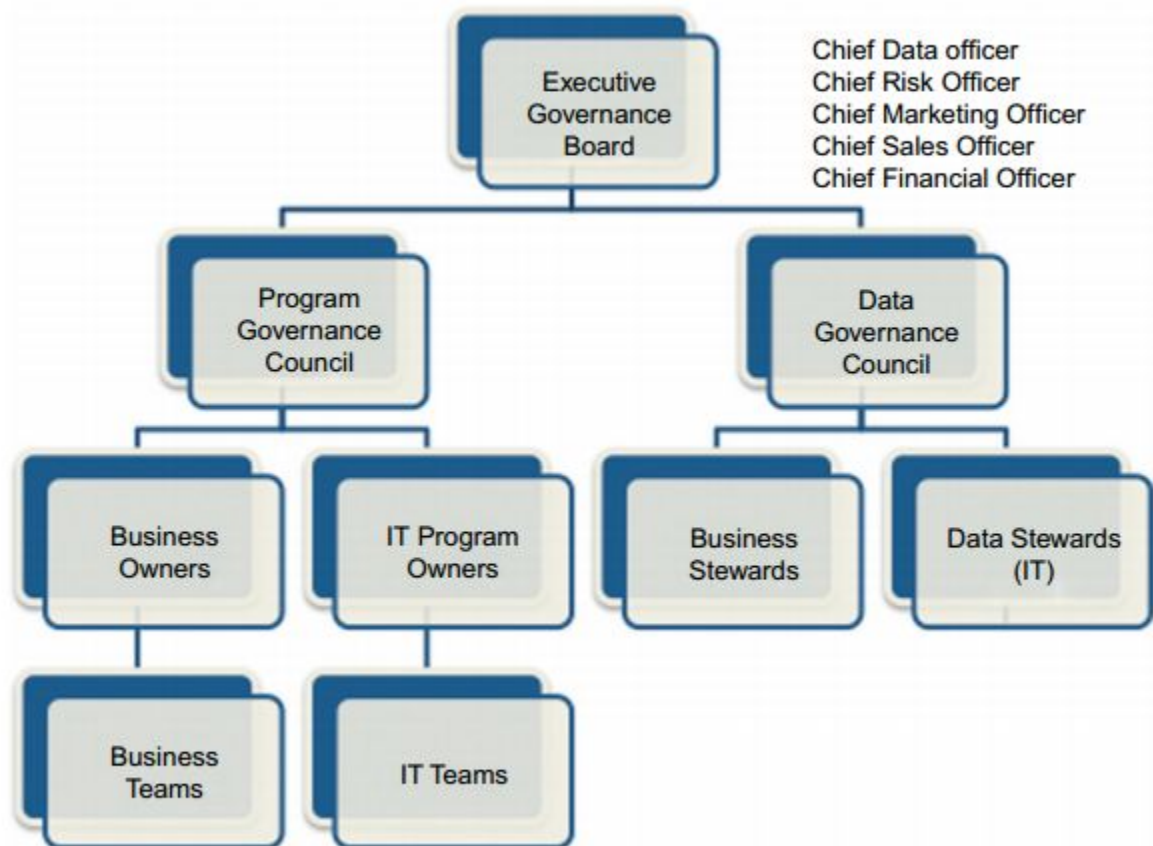


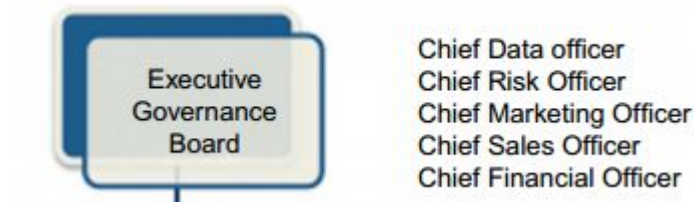
FIGURE 12.2

Data governance teams.

Conselho Executivo de Governança

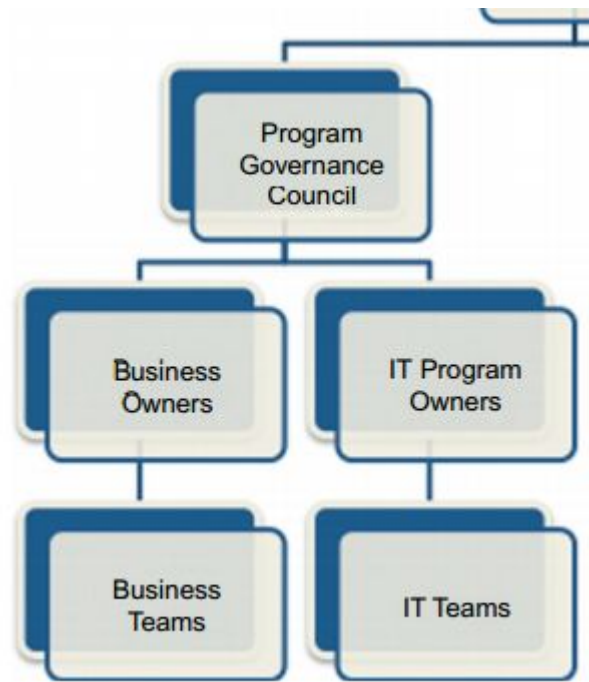
- Consiste em stakeholders das equipes executivas ou seus relatórios diretos.

- Responsável pela orientação geral e financiamento.



Conselho de governança do programa

- Consiste em proprietários de programas que são membros de nível **diretor** da organização executiva.
- Responsável pela direção geral do programa e pela:
- ◆ Gestão das equipes de TI
 - ◆ Coordenação das atividades
 - ◆ Gestão do orçamento
 - ◆ Priorização de tarefas e programas.



Conselho de governança de dados

Consiste em parceiros de negócios e de TI de cada unidade da empresa.

São responsáveis por tomar as decisões apropriadas para a integração dos dados na arquitetura corporativa, mantendo seus requisitos específicos dentro da mesma estrutura.

- Definição de dado
- Regras de qualidade de dados
- Metadados
- Política de acesso a dados
- Requisitos de criptografia
- Requisitos de ofuscação
- Políticas de gerenciamento de dados mestre
- Resolução de problemas e problema
- Políticas de retenção de dados

Governança: Considerações Finais

— — —

- Área importante para grandes e médias empresas para implementar uma iniciativa de transformação empresarial bem-sucedida
 - ◆ Gerenciamento de dados e informações

- Quando corretamente executada
 - ◆ Beneficia a empresa
 - ◆ Fornece confiança para as equipes de tomada de decisões
 - ◆ Reduz os riscos

Tecnologia

Ferramentas e tecnologias utilizadas na:

- aquisição de dados
- limpeza de dados,
- transformação de dados
- código de banco de dados

Ferramentas

— — —

- Qualidade dos dados
- Enriquecimento de dados
- Transformação de dados
- Arquivamento e retenção de dados
- Gerenciamento de dados mestre
- Metadados

Políticas de gerenciamento de informações

— — —

São políticas de:

- Aquisição
- Limpeza
- Transformação
- Retenção
- Segurança

Políticas de gerenciamento de informações

— — —

São políticas de:

→ **Aquisição**

- ◆ Aplicações em que as funções de entrada de dados são executadas
- ◆ Aplicativos Web e OLTP (Processamento de Transações em Tempo Real)
- ◆ Processos ETL ou CDC de data warehouse ou datamart
- ◆ Processos ETL de bancos de dados analíticos

Políticas de gerenciamento de informações

— — —

São políticas de:

- **Limpeza**
- **Transformação**

Transformar dados de origem em destino e incluem a transformação de níveis de granularidade, chaves, hierarquias, métricas e agregações.

Políticas de qualidade

Políticas de gerenciamento de informações

— — —

São políticas de:

→ **Retenção**

Dados online -> Dados Offline

Considerar os requisitos para dados que exigem suporte para gerenciamento de casos legais, gerenciamento de auditoria de conformidade e descoberta eletrônica.

O desafio na nova geração será sobre as técnicas mais eficientes de gerenciamento de dados.

Políticas de gerenciamento de informações

— — —

São políticas de:

→ **Segurança**

Criptografia de dados

Níveis de acesso de usuário.

Gerenciamento do ciclo de vida da informação para Big Data

— — —

- Muitos benefícios, mas também muitos desafios para a empresa
- As empresas estão implementando um projeto Big Data e projetando um programa de gerenciamento de ciclo de vida de informações para **ajudar a gerenciar** e **dinamizar** o Big Data
 - ◆ Aquisição e processamento
- Dados armazenados como arquivos (Hadoop ou NoSQL)
 - ◆ Simples exclusão de arquivos não é a solução certa
 - ◆ Preciso criar uma estratégia de retenção e arquivamento de dados robusta e bem definida

Dados de Mídias Sociais

Valor de tempo de vida dos dados
muito curto

vs

Processamento pode levar horas a
uma semana



Política de gestão do ciclo de vida da informação de Mídias Sociais

— — —

- Armazenamento de duas a quatro semanas para os dados brutos
- Seis a oito semanas para os conjuntos de dados processados após o processamento dos dados brutos
- Três anos para os dados resumidos agregados (usados para análises)

Questionamentos a se fazer

— — —

→ Governança de dados:

- ◆ Que dados devem ser retidos?
- ◆ Quais são os metadados e dados mestre associados que precisam ser processados com os dados a serem retidos?
- ◆ Que dados serão arquivados?

→ Segurança de dados:

- ◆ Quais são os requisitos de segurança do usuário e segurança de dados?

→ Metadados:

- ◆ Que taxonomias serão necessárias para gerenciar dados?

Benefícios da gestão do ciclo de vida da informação

— — —

- Aumenta a eficiência do processo
- Ajuda as empresas a otimizar a qualidade dos dados
- Acelera o ROI
- Estratégias de gerenciamento de dados melhoram os processos, incluindo:
 - ◆ Disponibilidade previsível do sistema
 - ◆ Otimização do desempenho do sistema
 - ◆ Maior reutilização dos recursos
 - ◆ Melhor gerenciamento de metadados e dados mestre
 - ◆ Otimizar gastos e custos
 - ◆ Reduzir os riscos relacionados com os dados

Análise de Big Data e Visualização de Dados



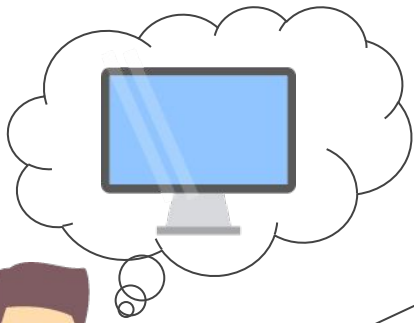
Maior impacto de Big Data é a capacidade de analisar eventos que aconteceram dentro e fora de uma organização e correlacioná-los para fornecer informações quase precisas sobre o que levou os resultados

Big Data Análise

“Combinação de técnicas tradicionais de análise e mineração de dados, juntamente com grandes volumes de dados, para criar uma plataforma fundamental para analisar, modelar e prever o comportamento de clientes, mercados, produtos, serviços e concorrência.”



João



Desktop Computing | Laptops & Notebooks | Servers & Storage | Hardware & Peripherals | Networking & Print | Servers & Peripherals | Servers & Peripherals | Tablets & E-Readers

The New MacBook Pro

State-of-the-art processors. All-new graphics. Breakthrough high-speed I/O.

LEARN MORE | PURCHASE

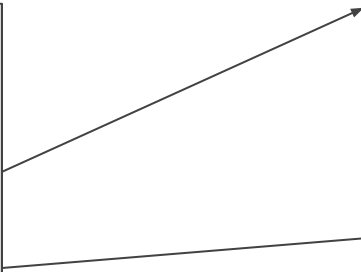
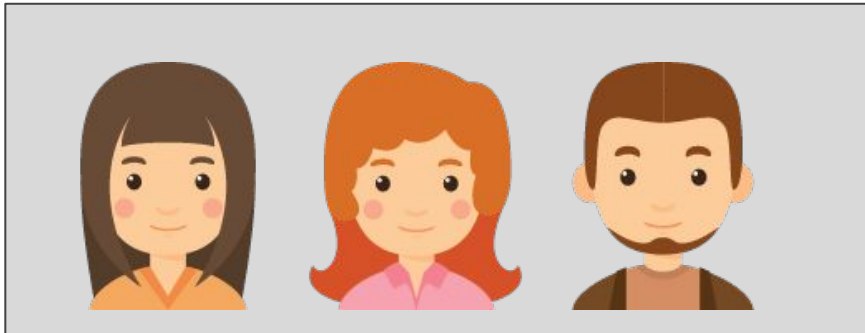
CATEGORIES

- Notebooks [SHOP NOW](#) FROM \$ 1,199
- Desktop Computers [SHOP NOW](#) FROM \$ 299
- Networking [SHOP NOW](#) FROM \$ 199
- Multimedia [SHOP NOW](#) FROM \$ 199
- Servers [SHOP NOW](#) FROM \$ 1,199

MANUFACTURERS

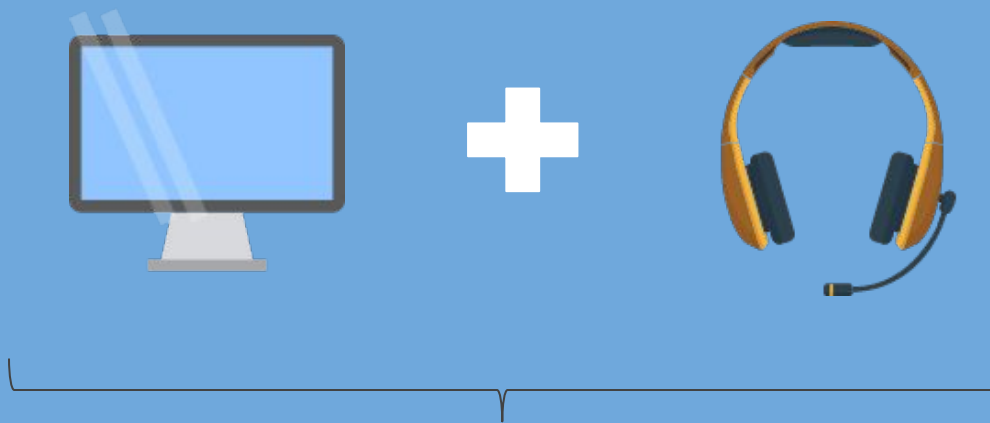
- Apple [SHOP NOW](#) FROM \$ 999
- Microsoft [SHOP NOW](#) FROM \$ 99
- Samsung [SHOP NOW](#) FROM \$ 99
- Alienware [SHOP NOW](#) FROM \$ 1,199
- Panasonic [SHOP NOW](#) FROM \$ 1,199

[Specials](#) [Advanced Search](#) [Reviews](#) [Create an Account](#) [Log In](#) [Shipping & Returns](#) [Contact Us](#)





- Últimas compras de João.
- Últimas pesquisas
- Perfil
- Perfil dos compradores..



João, comprando também esse headset você ganha 40% de desconto!

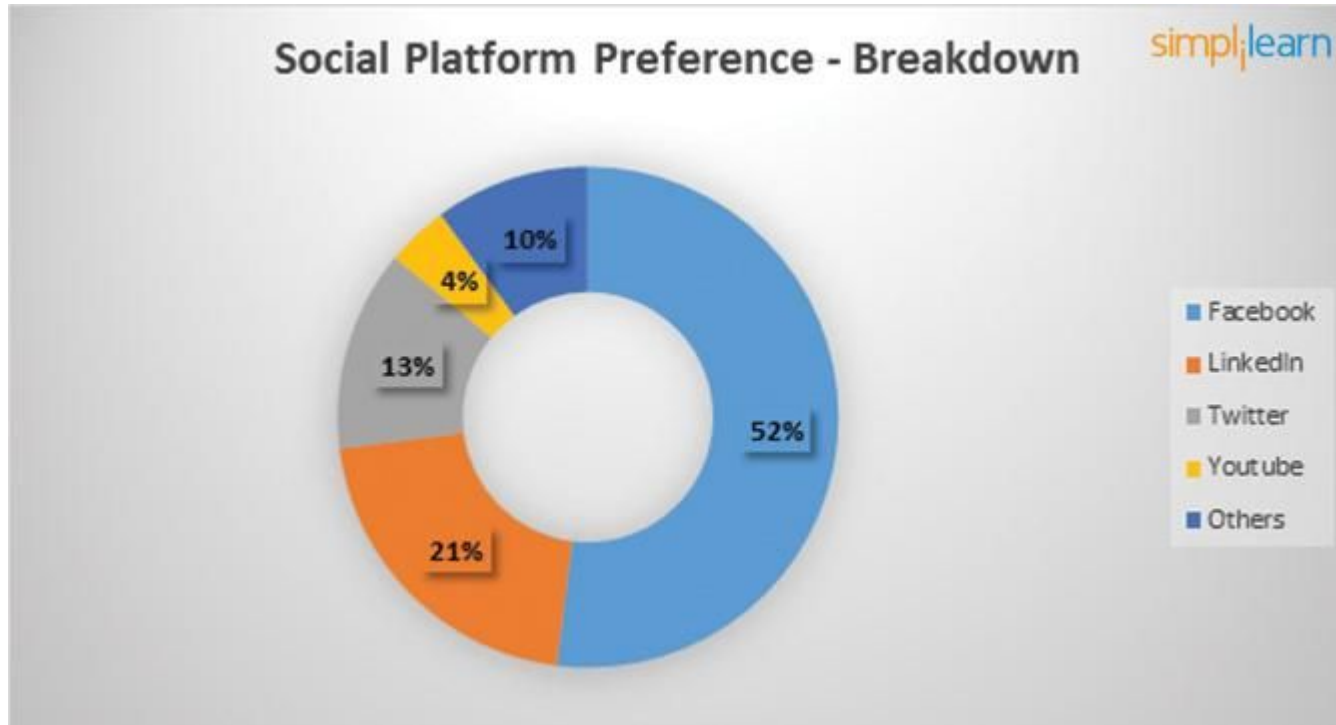


Descoberta de dados

— — —

- Aquisição de dados
 - ◆ Processo de coleta de dados para descoberta.
- Etiquetagem de dados
 - ◆ Processo de criação de um link de identificação nos dados para a integração de metadados
- Classificação de dados
 - ◆ Processo de criação de subconjuntos de pares de valores para processamento e integração de dados
- A modelagem de dados
 - ◆ Processo de criação de um modelo para visualização ou análise de dados

Facebook



plataforma social #1 para os comerciantes

Facebook

— — —

Todos os dias ele é alimentado por:

- 10 bilhões de mensagens
- 4,5 mil milhões de clicks no botão 'like'
- 350 milhões de envios de imagem

Facebook sabe quem são nossos amigos, nossa aparência, onde estamos, o que estamos fazendo, nossos gostos, nossos desgostos, e muito mais.

Facebook

— — —

→ Cookies de rastreamento:

- ◆ Se um usuário estiver conectado ao Facebook e ao mesmo tempo navega na Web, o Facebook pode acompanhar os sites que eles visitam

→ Reconhecimento Facial:

- ◆ Rastreia seus usuários em todo os outros perfis do Facebook Web, devido aos dados de imagem que é armazenado com eles quando o usuário compartilha

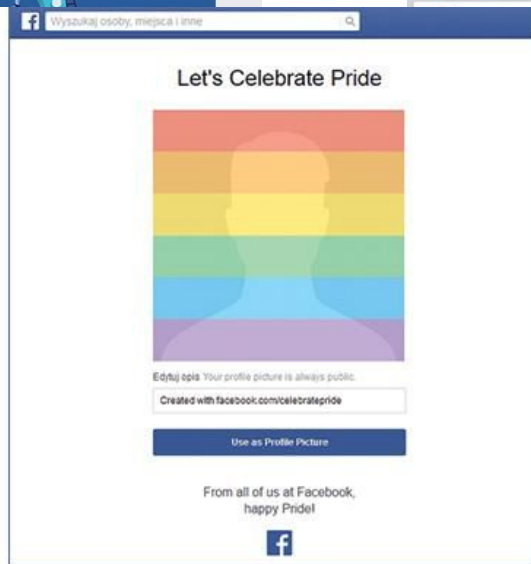
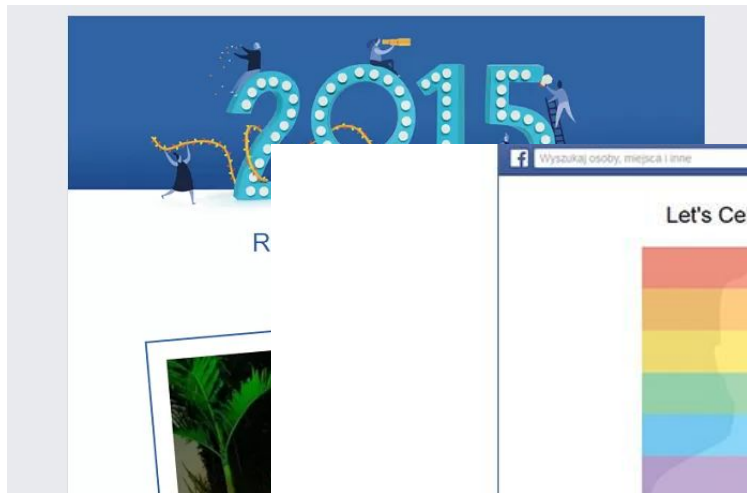
→ Sugestão Tag :

- ◆ Sugestões Tag podem ser feitas em fotos do usuário devido ao processamento de imagem e reconhecimento facial

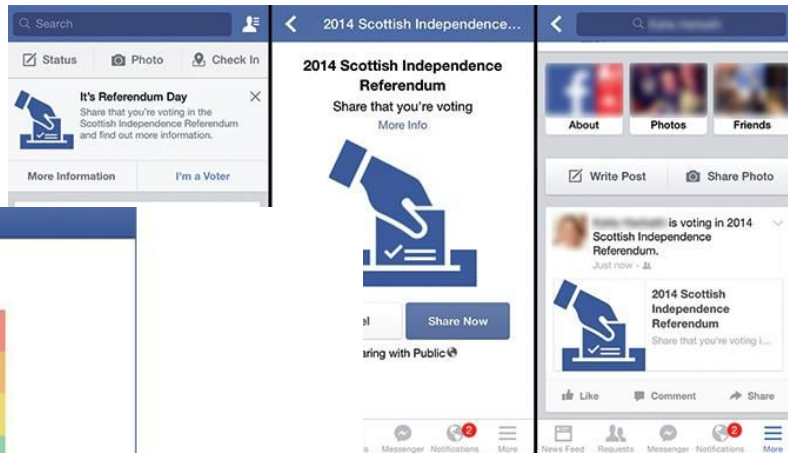
→ Análise do “like”:

- ◆ Pesquisa realizada pela Universidade de Cambridge e Microsoft Research
- ◆ Podem prever com bastante precisão a sua **orientação sexual, satisfação com a vida, inteligência, estabilidade emocional, religião, uso de álcool e uso de drogas, status de**

Como o facebook usa o big data?



Data credits: pablik.pl



Data credits: clickz.com

Facebook vs Google

- As informações do Google sobre quem somos é muitas vezes uma "melhor estimativa" com base em quais sites que visitamos e quais termos de pesquisa que estamos usando
- O Facebook nos pede explicitamente quem somos, onde vivemos e aquilo que está interessado.

Facebook cresceu 129% entre 2011 e 2013

49% no Google

Visualização de Dados

— — —

- Procuram por padrões, tendências e correlações
- Dados transformados em imagens gráficas, a identificação de padrões e tendências torna-se mais eficaz
- Dados contam uma história sobre o fenômeno estudado e, assim, as soluções parecem saltar diante dos olhos dos analistas.

Tableau Gallery

tableau public

GALLERY

AUTHORS

BLOG

RESOURCES

ACTIVITY

SIGN IN



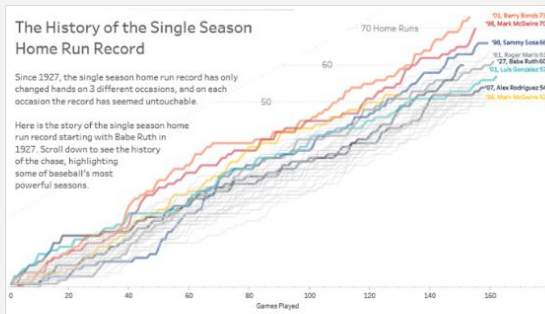
Gallery / Greatest Hits

Stunning data visualization examples from across the web created with Tableau Public.

Greatest Hits

Viz of the Day

All



The History of the Single Season Home Run Record

Our 3rd 2016 Iron Viz Feeder Winner looks at Home Run Records.

Record Breaking Coasters

<https://public.tableau.com/en-us/s/gallery>

Power BI Gallery


Microsoft | Power BI | Products | Solutions | Partners | Learn

WELCOME TO THE DATA STORIES GALLERY

A showcase of data storytelling brought to life.


Featured | Top Kudos | Recently posted | All messages

Options | Submit your data story




Power BI "See 'n Say"

deldersveld | Featured | 8



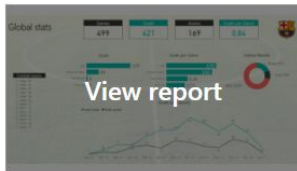
ROAD ACCIDENTS ANALYSIS 2006-2016 VICTORIA AUSTRALIA

BhaveshPatel | Featured | 6




Ohio Primary and Secondary Education Performance

smoupre | Featured | 6





Incredible Lionel Messi

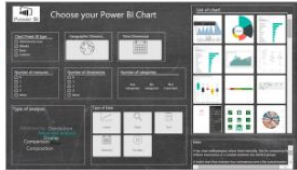
Excelside | Featured | 4



IMDB - Rotten Tomatoes Film Analysis







Choose your Power BI Chart



As of February 2016

Conclusão

— — —

- Big Data é um ambiente muito promissor que pode agregar infinitos valores a uma organização, se aplicado e gerenciado da forma correta.
- Políticas de gerenciamento do ciclo de informação bem definidas, para amenizar ao máximo problemas de gestão de dados que venham emergir no futuro.
- Ferramentas cada dia mais práticas e acessíveis, gerar visualizações de dados em cima de big data está a cada dia mais fácil

Referências

KRISHNAN, Krish. Data Warehousing in the Age of Big Data. 1 ed. Waltham, MA, USA: Elsevier Inc, 2013.

CHEN, Min; MAO, Shiwen; LIU; YUNHAO. Big Data: A Survey.

SOUSA, Flaviano. Information Lifecycle Management, 2006.

TESSAROLO, Pedro. MAGALHAES, Willian. A era do Big Data no conteúdo digital: Os dados estruturados e não estruturados.

ZICARI, Roberto. Big Data: Challenges and Opportunities.

Referências

How Facebook is Using Big Data - The Good, the Bad, and the Ugly. Disponível em:
<<https://www.simplilearn.com/how-facebook-is-using-big-data-article>>

Big Data Analytics: Você sabe o que é? Disponível em:
<<http://www.bigdatabusiness.com.br/voce-sabe-o-que-e-big-data-analytics/>>

O que é Power BI? Disponível em:
<<https://powerbi.microsoft.com/pt-br/what-is-power-bi/>>

Tableau Public. Disponível em:
<<http://www2.serg.inf.puc-rio.br/docs/TableauGuides/TableauPublic-TranslationPT-BR.pdf>>

Dúvidas?



Obrigada!