



BIG DATA

O NOVO DESAFIO DAS EMPRESAS E PROFISSIONAIS DO MERCADO

ROSANGELA MARQUESONE

Tópicos



O que é Big Data

Desafios para empresas e profissionais

Para inspirar

Tendências futuras

Como você define **Big Data**?

Definição de Big Data

Big Data faz referência ao grande volume, variedade e velocidade de dados que demandam formas inovadoras e rentáveis de processamento da informação, para melhor percepção e tomada de decisão.

Gartner

Definição de Big Data

Big Data faz referência ao grande volume, variedade e velocidade de dados que demandam formas inovadoras e rentáveis de processamento da informação, para melhor percepção e tomada de decisão.

Gartner

VOLUME

VARIEDADE

VELOCIDADE

Os 3 Vs de Big Data

VOLUME

40 milhões
de artigos

Wikipedia

1.23 bi
usuários
ativos/dia

Facebook

80 milhões
de fotos por
dia

Instagram

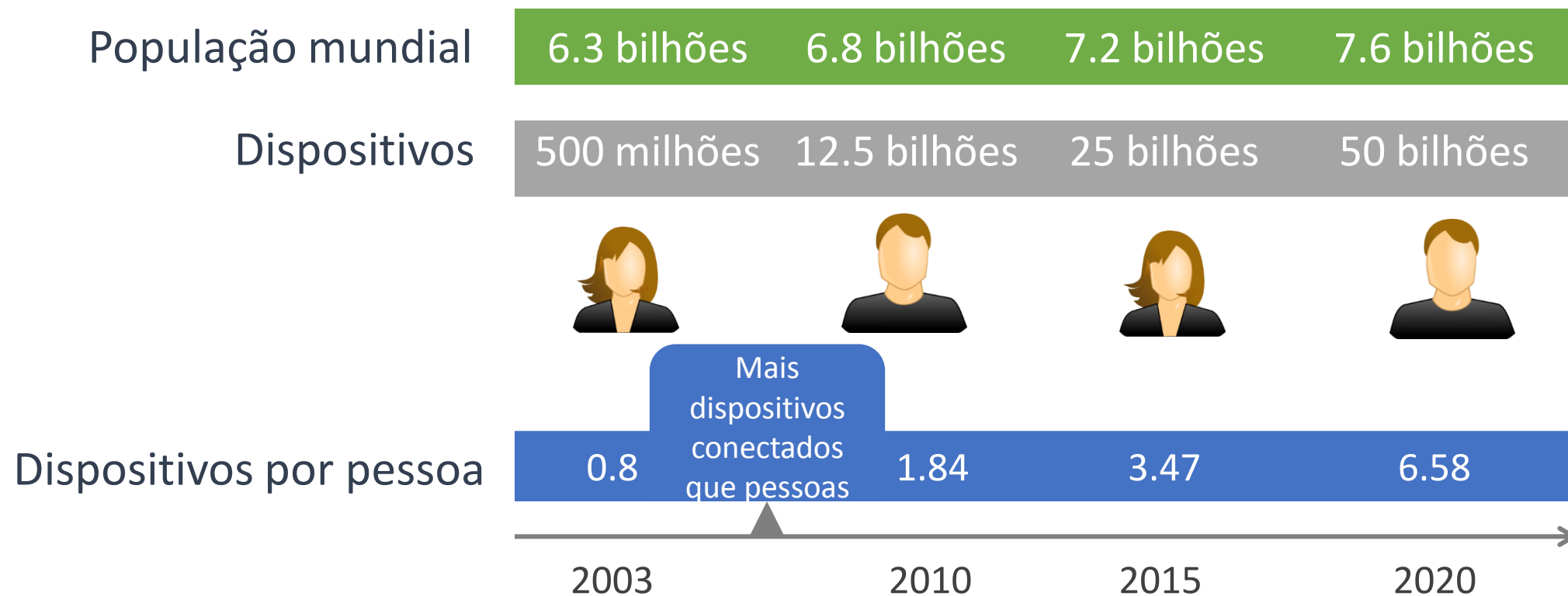
1 bilhão
usuários

Whatsapp

4 bilhões
Visualizações
por dia

Youtube

VOLUME



VARIEDADE

80%

dos dados globais atualmente são
não-estruturados

CENÁRIO TRADICIONAL

Dados
transacionais



VARIEDADE

CENÁRIO ATUAL

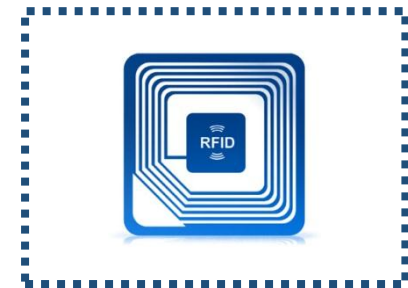
VARIEDADE



Dados
transacionais



Repositório
de dados

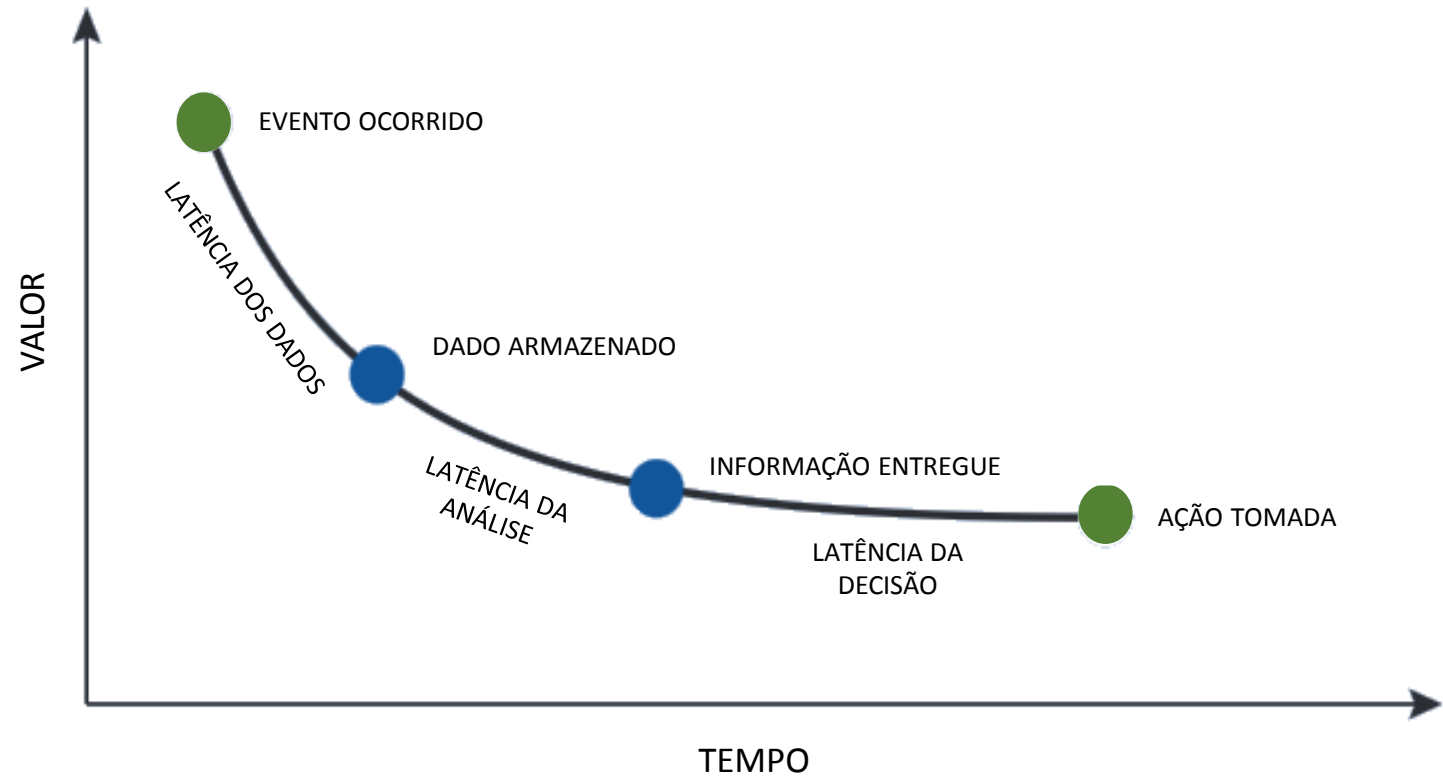


O que acontece em 30 segundos na Internet?



VELOCIDADE

O valor dos dados é reduzido com o passar do tempo



VELOCIDADE

Definição de Big Data

Big Data faz referência ao grande volume, variedade e velocidade de dados **que demandam formas inovadoras e rentáveis de processamento da informação,** para melhor percepção e tomada de decisão.

Gartner

Tópicos

O que é Big Data

Desafios para empresas e profissionais

Para inspirar

Tendências futuras

Desafio

Inovação no
armazenamento
dos dados



REDES SOCIAIS



VÍDEO



ÁUDIO



EMAIL



MENSAGEM



**ONDE ARMAZENAR
ESSES DADOS?**

**COMO ESTRUTURAR
ESSES DADOS?**

**COMO CONSULTAR
ESSES DADOS?**



TRANSAÇÕES



MOBILE



IMAGEM



DOCUMENTOS



GEOLOCALIZAÇÃO

Tecnologias para armazenamento de dados

NoSQL

=

Not Only SQL

*“Conjunto de conceitos que permite o processamento rápido e eficiente de conjuntos de dados com foco em **desempenho, confiabilidade e agilidade**”.*

Making sensing of NoSQL

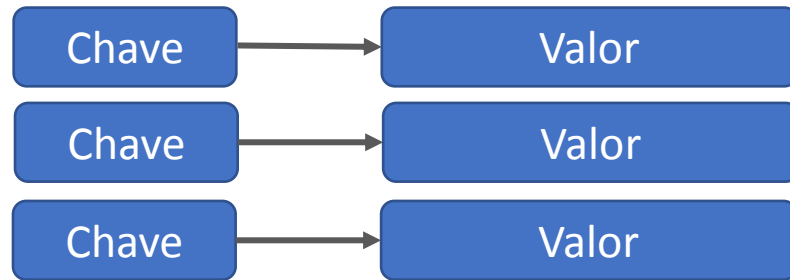
NoSQL

Diferentes formas de armazenamento



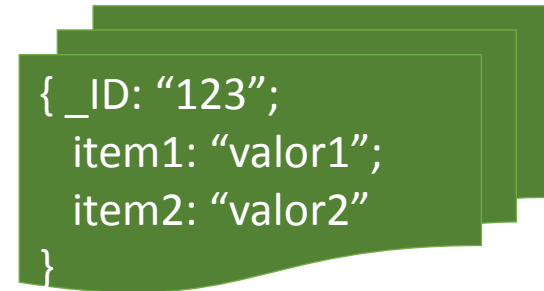
NoSQL

Orientado a chave-valor



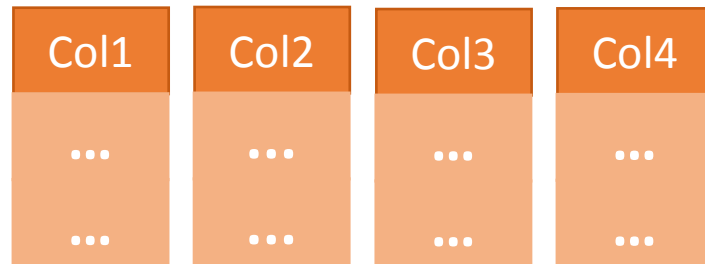
“O mais simples”

Orientado a documentos



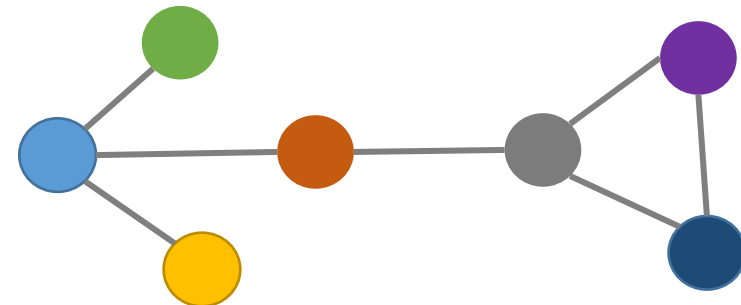
“O mais popular”

Orientado a colunas



“O mais complexo”

Orientado a grafos



“O mais especializado”

Desafio

Inovação no
processamento
dos dados

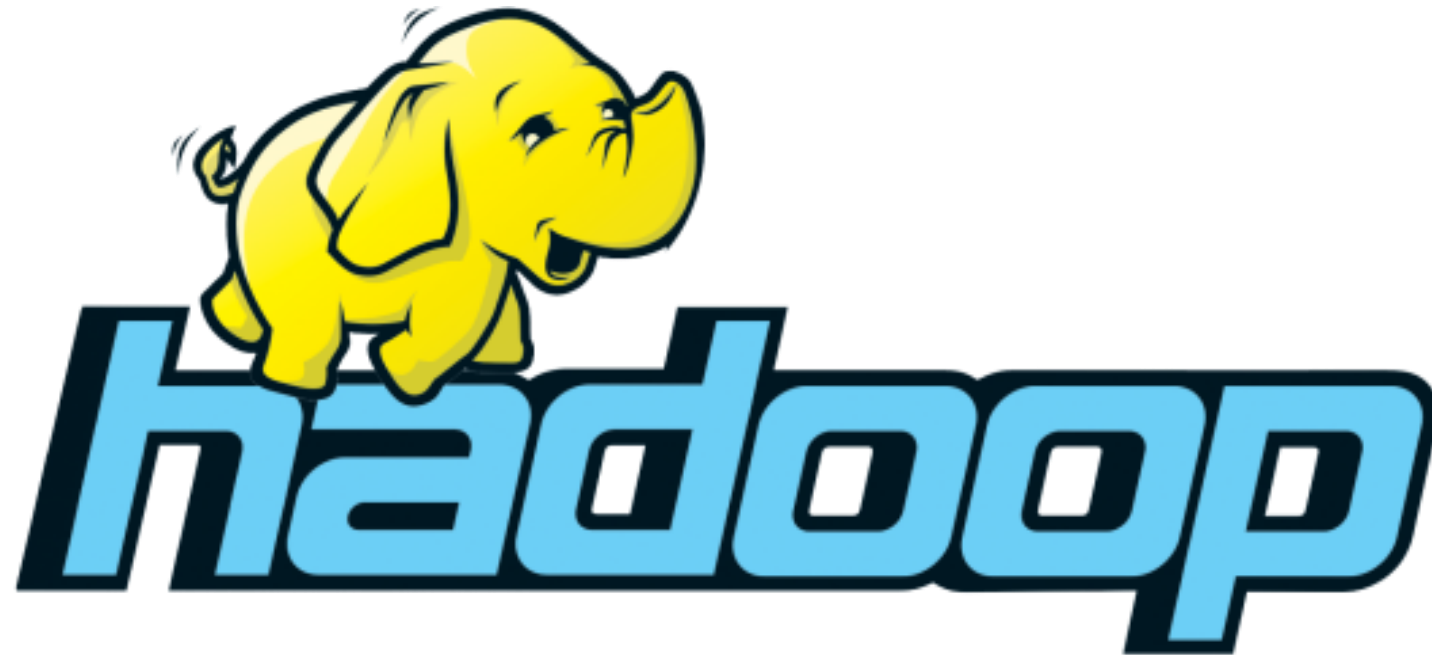


TECNOLOGIAS DE BIG DATA

=

Novas tecnologias capazes de oferecer escalabilidade, disponibilidade, flexibilidade e desempenho para a manipulação de grande volume de dados

Inovação no processamento dos dados



Hadoop

Solução **open source** que permite a execução de aplicações de Big Data utilizando milhares de máquinas

Oferece recursos de armazenamento, gerenciamento e processamento **distribuído** de dados (SO de Big Data)

Benefícios:

Redução de custo

Escalabilidade

Flexibilidade

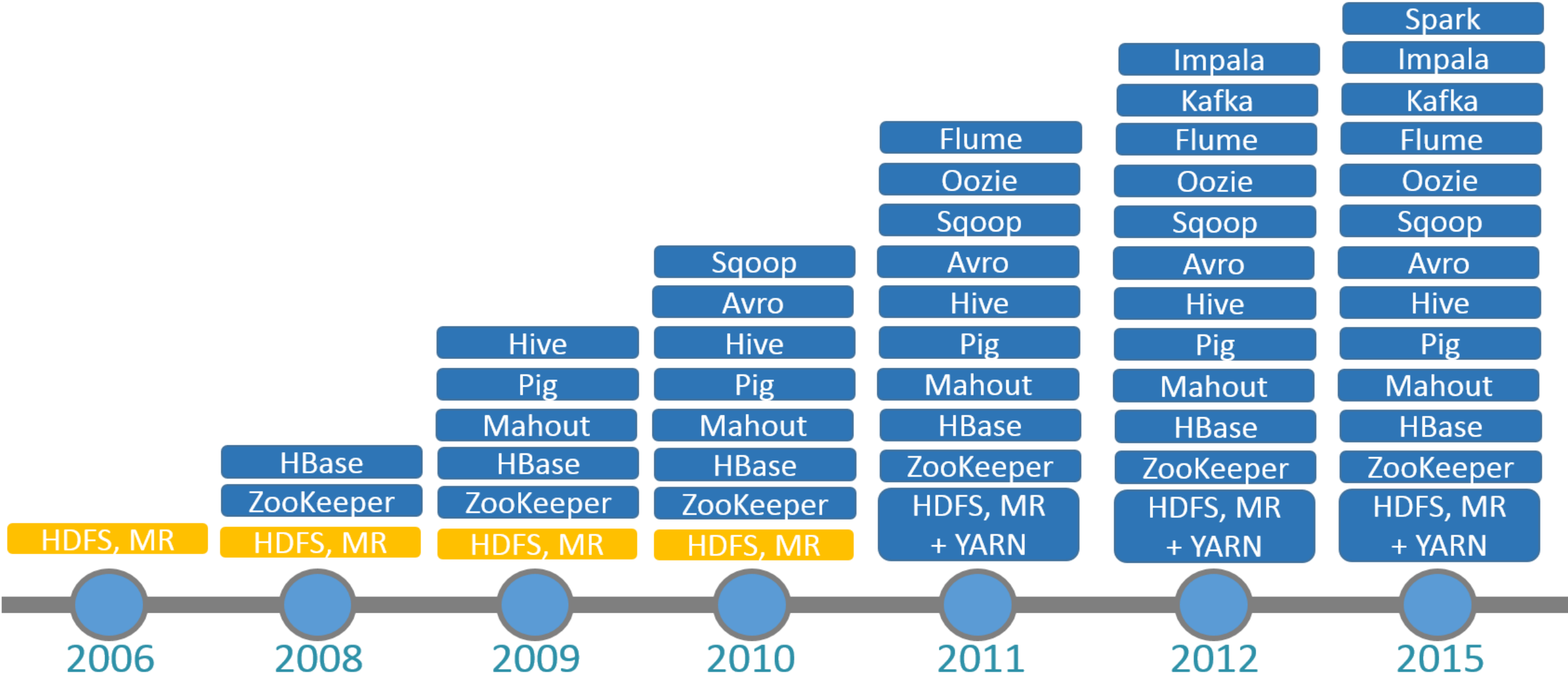
Desempenho



Hadoop - Quem utiliza



Ecosystem Hadoop



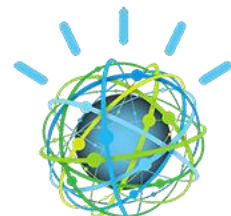
Demais tecnologias



python™



elasticsearch.



IBM Watson



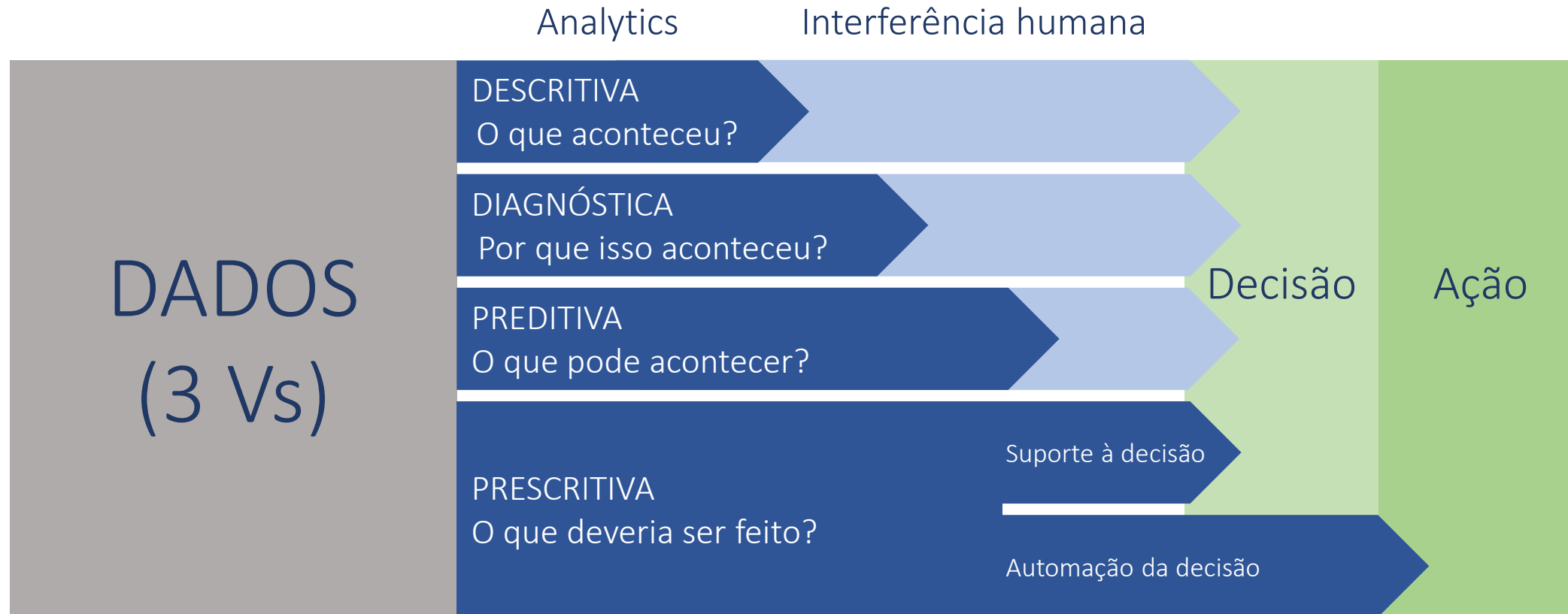
Desafio

Inovação na
análise dos
dados

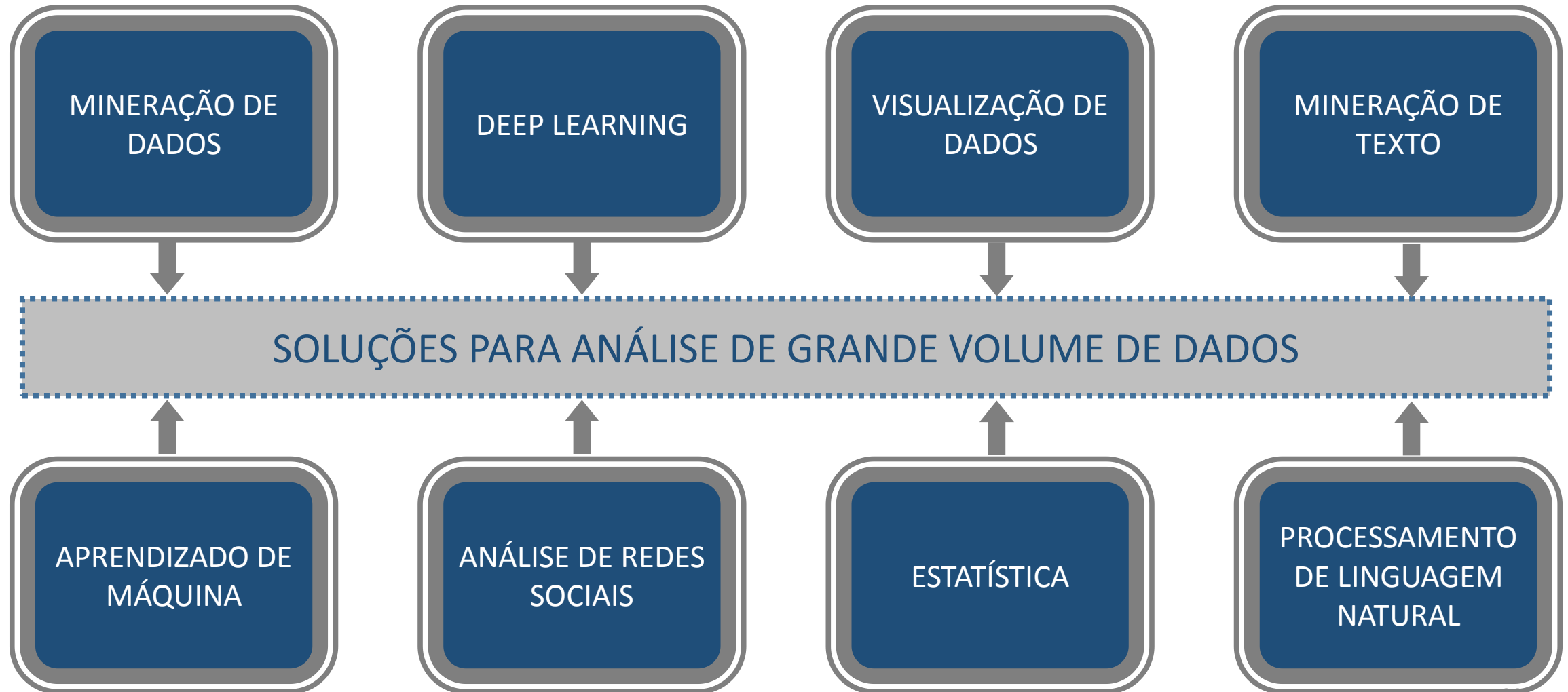


© pixelart.com

Big Data Analytics

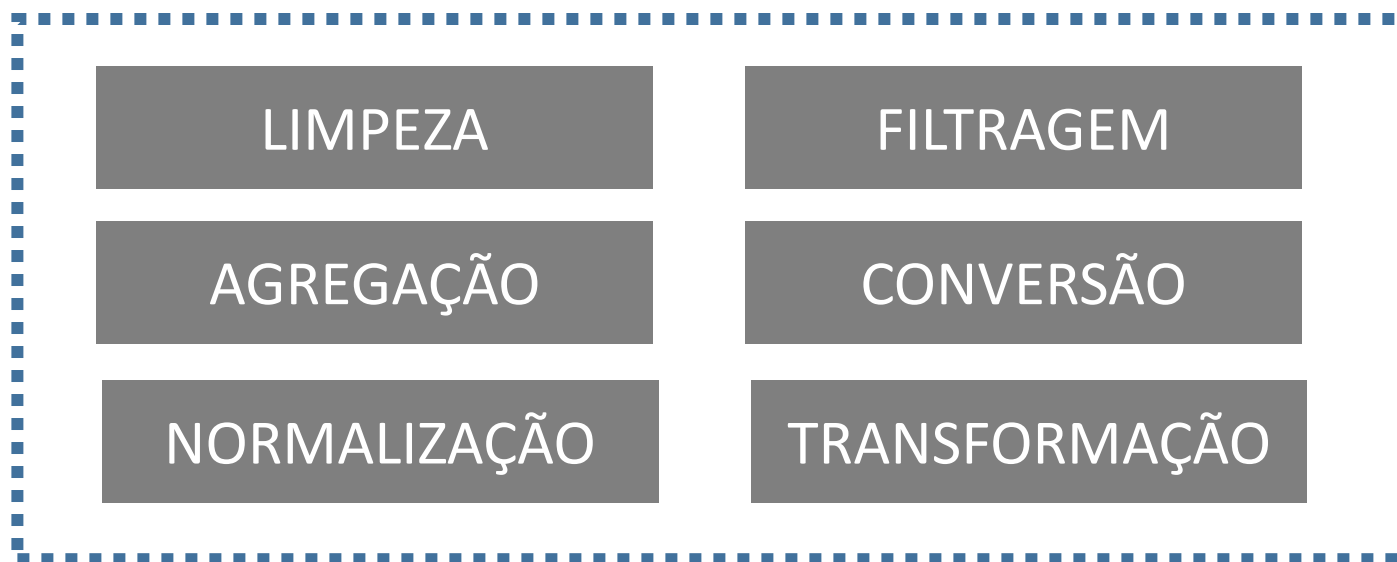


Inovação na análise de dados



80%

do processo de análise é gasto preparando os dados



CAPTURA

PREPARAÇÃO

ANÁLISE

Desafio

Encontrar
colaboradores
capacitados



Cientista de dados

Mercado de Big Data: carência de profissionais qualificados

www.broadtec.com.br/mercado-de-big-data-carencia-de-profissionais-qu... ▼

7 de abr de 2015 - Estimativas do Gartner apontam que este ano serão criados 4,4

Avanço de Big Data no Brasil deve esbarrar em falta de mão-de-obra

Por Rafael Romer  | 10.02.2014 às 12h11

Cientista de dados: o profissional mais cobiçado do século 21 ...

hbrbr.uol.com.br/cientista-de-dados-o-profissional-mais-cobicado-do-seculo-21/ ▼

Goldman é um bom exemplo de um novo ator de destaque em organizações: o "cientista de dados", um profissional de alto escalão com qualificação e ...

Brasil sofre com a falta de profissionais para o big data

Ana Paula Lobo e Pedro Costa ... 12/05/2015 ... Convergência Digital

26/12/2014

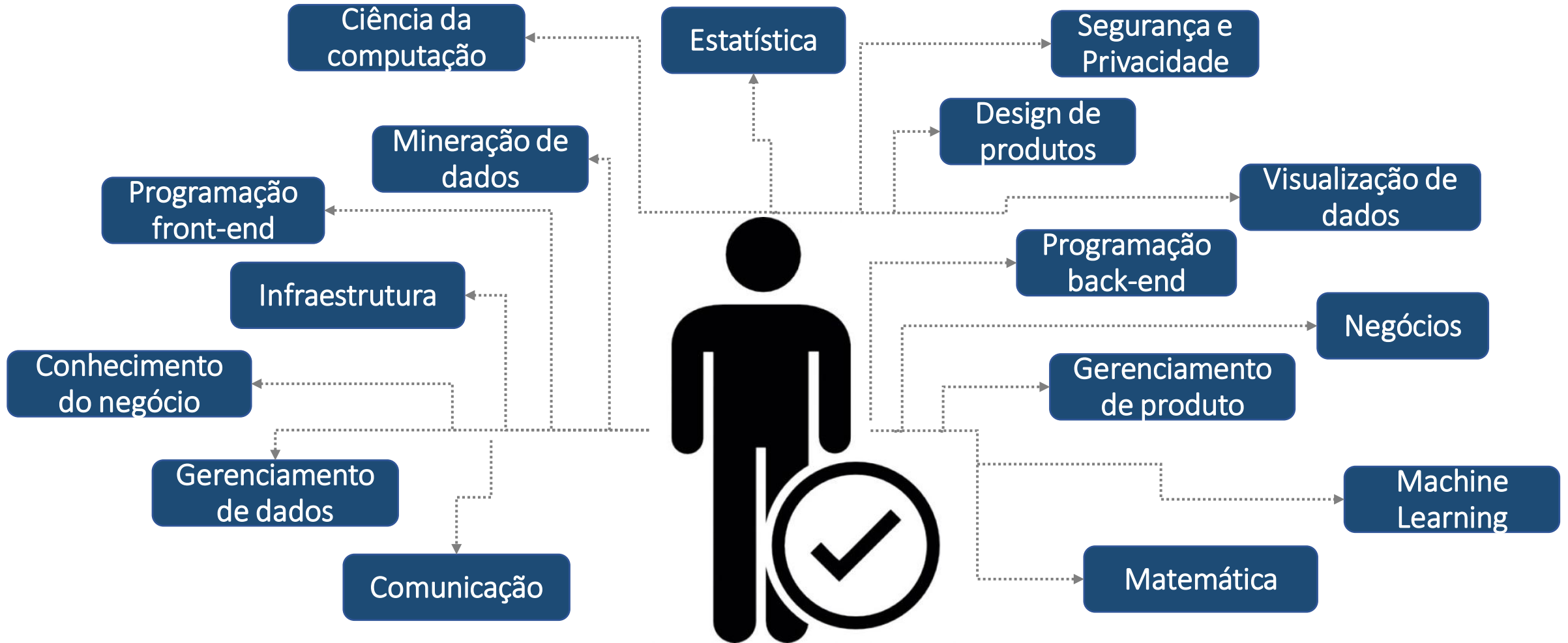
Falta de mão de obra especializada é principal entrave para Big Data

Especialista sugere treinar profissionais internos para acelerar processo de qualificação da mão de obra

Cientista de dados

Quais habilidades as empresas estão buscando em um **cientista de dados**?

Cientista de dados



Cientista de dados

PROFISSÃO DO ANO DE 2016: CIENTISTA DE DADOS

salário médio anual: US\$ 128.240

Eleita a **profissão do ano de 2016** pelo site Americano de empregos CareerCast.com, considerando critérios como: ambiente de trabalho, renda, nível de stress e perspectiva de contratação

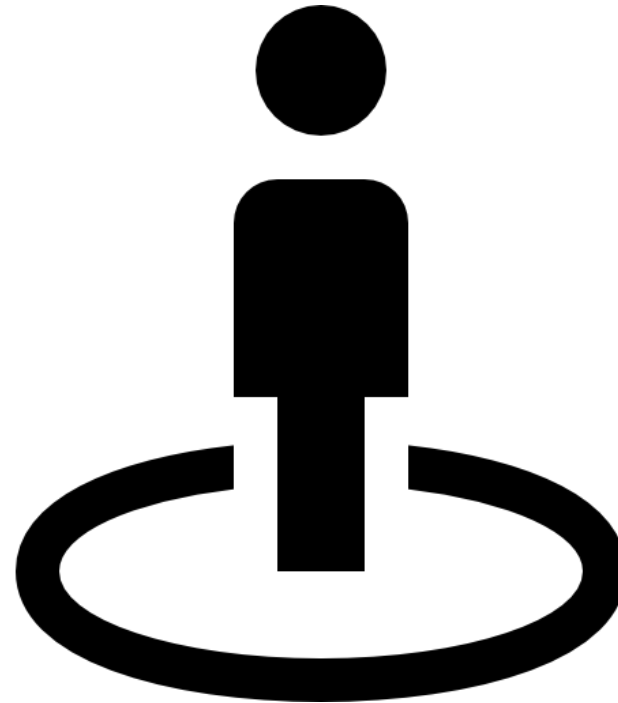
Profissão nº 2: estatístico
salário médio anual: US\$ 79.990



Fonte: <http://g1.globo.com/economia/concursos-e-emprego/noticia/2016/05/veja-lista-com-10-melhores-e-piores-profissoes-para-2016.html>

Cientista de dados

UM PROJETO DE BIG DATA NÃO SE FAZ SOZINHO!



Cientista de dados

Um projeto de Big Data deve ser formado por uma **equipe multidisciplinar** altamente **qualificada**



“Data Science is a team sport”

Barack Obama

Definição de Big Data

Big Data faz referência ao grande volume, variedade e velocidade de dados que demandam formas inovadoras e rentáveis de processamento da informação,
para melhor percepção e tomada de decisão.

Gartner

Tópicos

O que é Big Data

Desafios para empresas e profissionais

Para inspirar

Tendências futuras



TARGET

Objetivo: analisar hábitos de compras e **prever a gravidez de clientes**

Solução: Algoritmos de mineração de dados para análise preditiva de dados

Padrões identificados:

- Compras de loção sem cheiro: **segundo trimestre**
- Suplementos de cálcio e zinco: **primeiras 20 semanas**
- Sabonete e pacotes de algodão: **está para nascer**
- 25 produtos analisados, ao todo

Resultado: Campanha de marketing personalizado

Analytics



PIRELLI

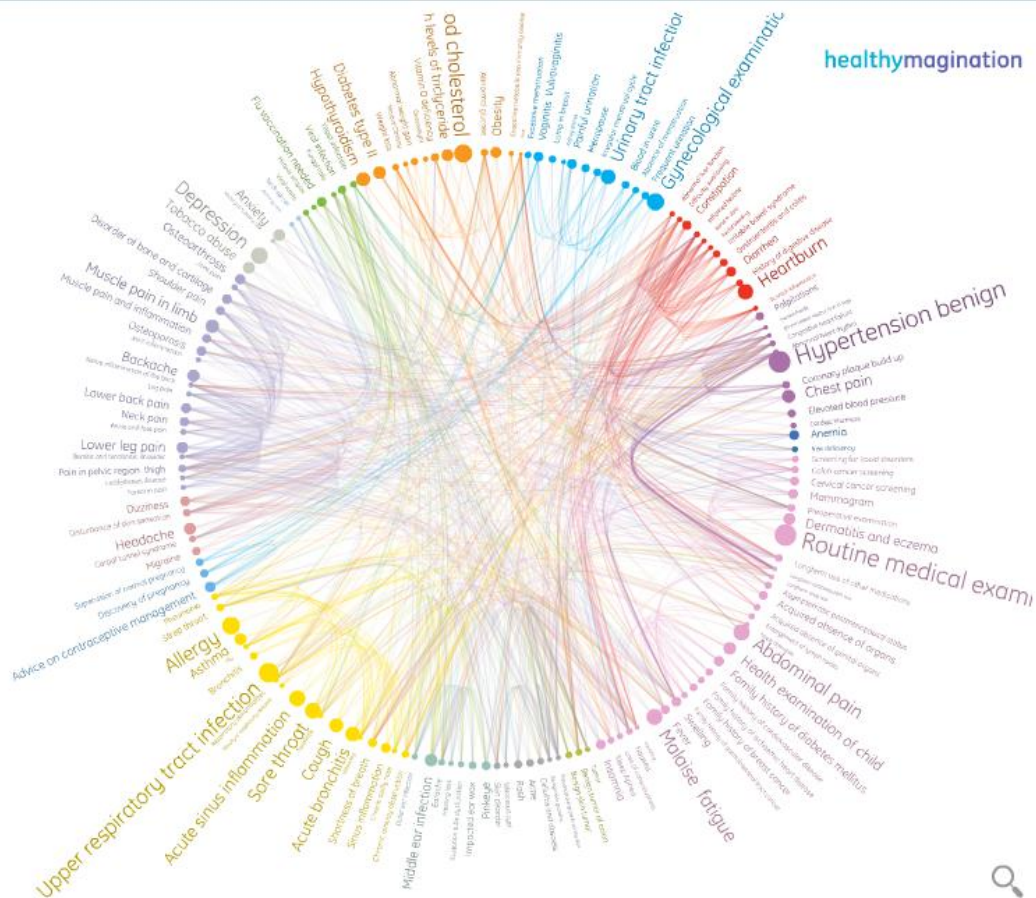
Objetivo: acelerar a coleta de dados e garantir que as informações necessárias para operar suas operações de forma eficiente estivessem disponíveis para o negócio em **tempo real**.

Solução: Adoção de tecnologias de processamento em memória de grande volume de dados.

Resultado: os relatórios de vendas e distribuição funcionaram **8.4 vezes mais rápido** do que antes. O acesso mais rápido à informação do produto, por sua vez, possibilitou uma tomada de decisão mais informada para garantir que seus produtos alcançassem seus clientes a tempo.

Processamento em tempo real

- Categories
- Blood Diseases
 - Circulatory System
 - Digestive System
 - Genitourinary System
 - Hormone Nutrition Immunity
 - Infections
 - Injury and Poisoning
 - Mental Health
 - Musculoskeletal System
 - Nervous System
 - Pregnancy Early Development
 - Respiratory System
 - Sensory Organs
 - Skin Conditions
 - Tumors
 - Unclassified



GENERAL ELETRIC

Objetivo: visualizações do relacionamento entre sintomas de doenças

Solução: analisar registros de 7.2 milhões de pacientes capturados nos equipamentos da GE

Resultado: visualização de dados interativo. Apresenta categoria dos sintomas, como os sintomas estão relacionados e prevalência entre homens e mulheres

Mineração de dados

Visualização de dados



EMPRESAS COM NEGÓCIOS CENTRADOS EM
DADOS





TOMA DECISÕES POR
MEIO DOS DADOS



TRATA OS DADOS COMO
UM ATIVO DA EMPRESA



EXPLORA OS DADOS
CAPTURADOS

E SUA EMPRESA, É GUIADA POR DADOS?

FACILITA O ACESSO AOS
DADOS



DEFINE O QUE QUER
MEDIR



OFERECE FERRAMENTAS
ANALÍTICAS



Tópicos

O que é Big Data

Desafios para empresas e profissionais

Para inspirar

▶ **Tendências futuras**

O que se falava antes...

“Big Data pode oferecer vantagem competitiva para as empresas”

O que se fala agora...

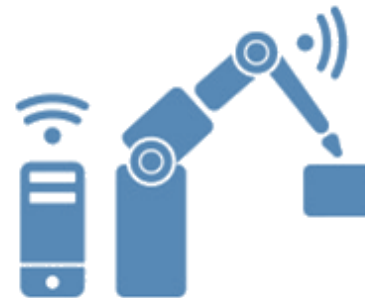
“Big Data é necessário para que as empresas se **mantenham competitivas**”

O QUE ESPERAR PARA OS PRÓXIMOS ANOS EM BIG DATA?

CHAT BOTS



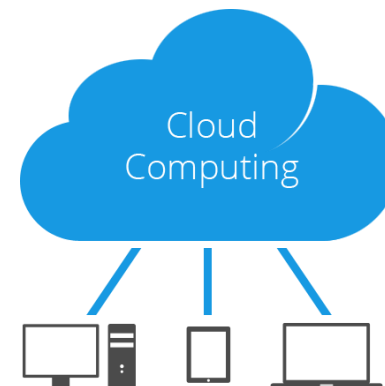
INDÚSTRIA 4.0



SEGURANÇA DOS DADOS



**CARROS
AUTÔNOMOS**



**COMPUTAÇÃO EM
NUVEM**



**INTEGRAÇÃO DOS
DADOS**

Considerações finais

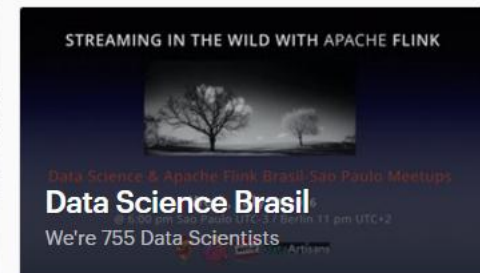
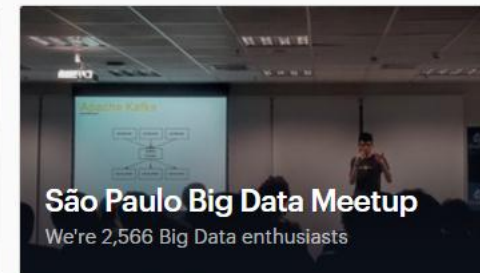
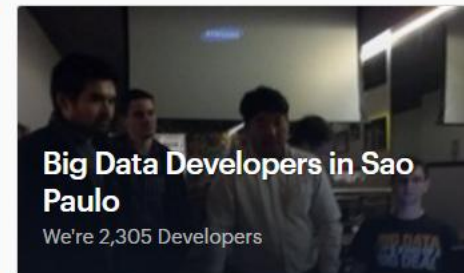
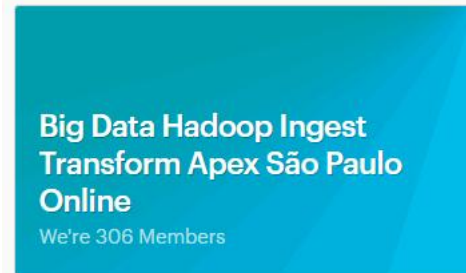
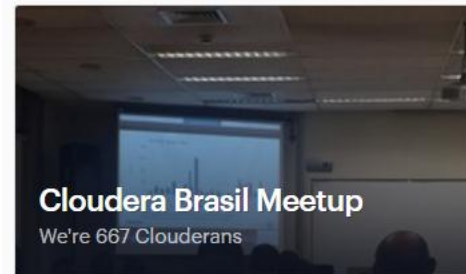
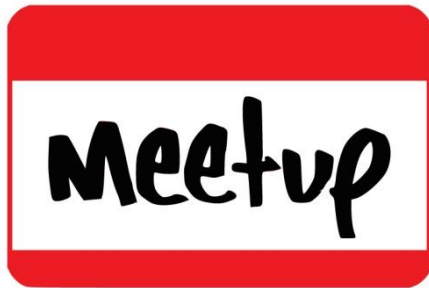
DADO É O NOVO PETRÓLEO!

Precisamos encontrá-lo, extraí-lo,
refiná-lo, distribuí-lo e monetizá-lo!

David Buckingham

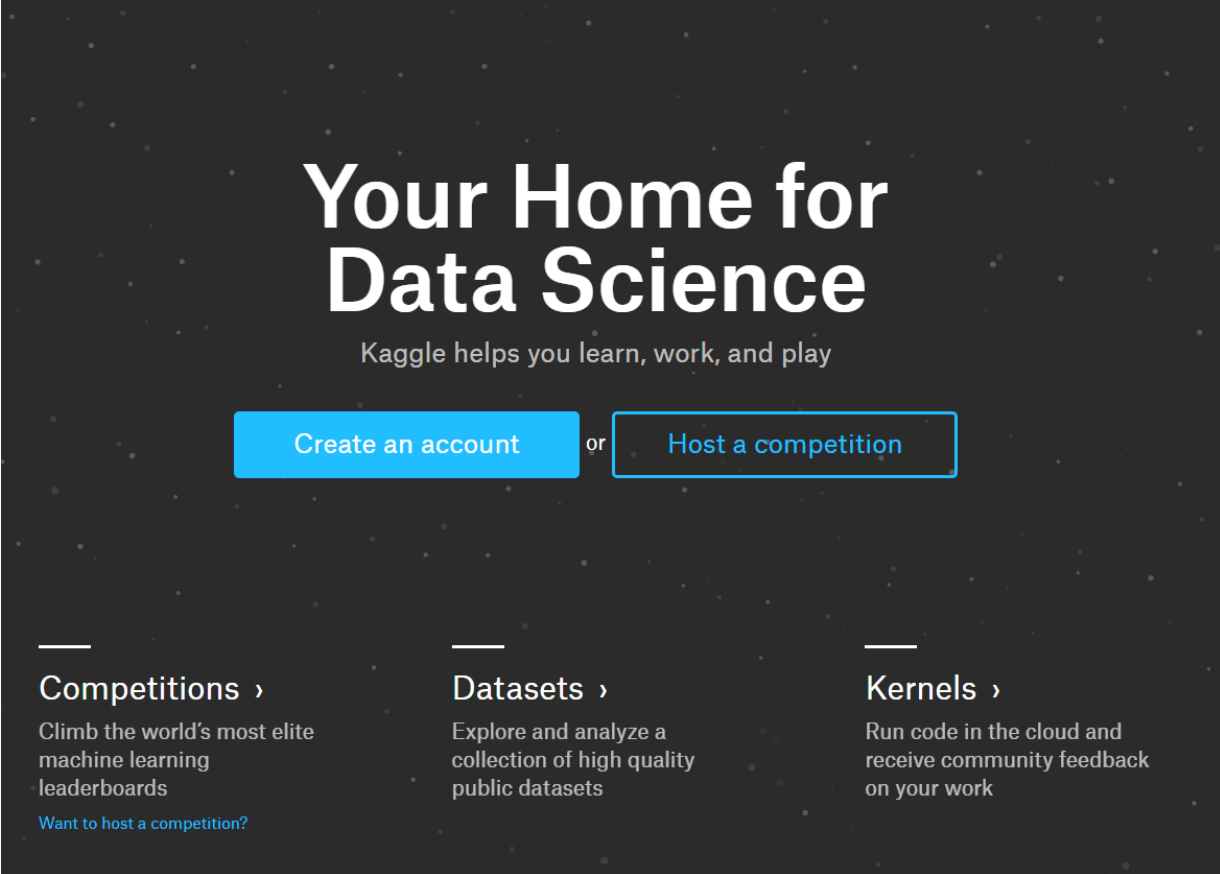


Para saber mais...



Para saber mais...

kaggle

A screenshot of the Kaggle homepage. The background is dark with a starry space pattern. At the top center, the text "Your Home for Data Science" is written in large white font. Below it, in a smaller white font, is the tagline "Kaggle helps you learn, work, and play". Two blue buttons are positioned below the tagline: "Create an account" on the left and "Host a competition" on the right, with the word "or" in small white text between them. At the bottom, there are three columns of text, each preceded by a horizontal line. The first column is titled "Competitions" and describes climbing world's most elite machine learning leaderboards, with a link "Want to host a competition?". The second column is titled "Datasets" and describes exploring and analyzing a collection of high quality public datasets. The third column is titled "Kernels" and describes running code in the cloud and receiving community feedback on work.

Your Home for Data Science
Kaggle helps you learn, work, and play

[Create an account](#) or [Host a competition](#)

Competitions ›
Climb the world's most elite machine learning leaderboards
[Want to host a competition?](#)

Datasets ›
Explore and analyze a collection of high quality public datasets

Kernels ›
Run code in the cloud and receive community feedback on your work

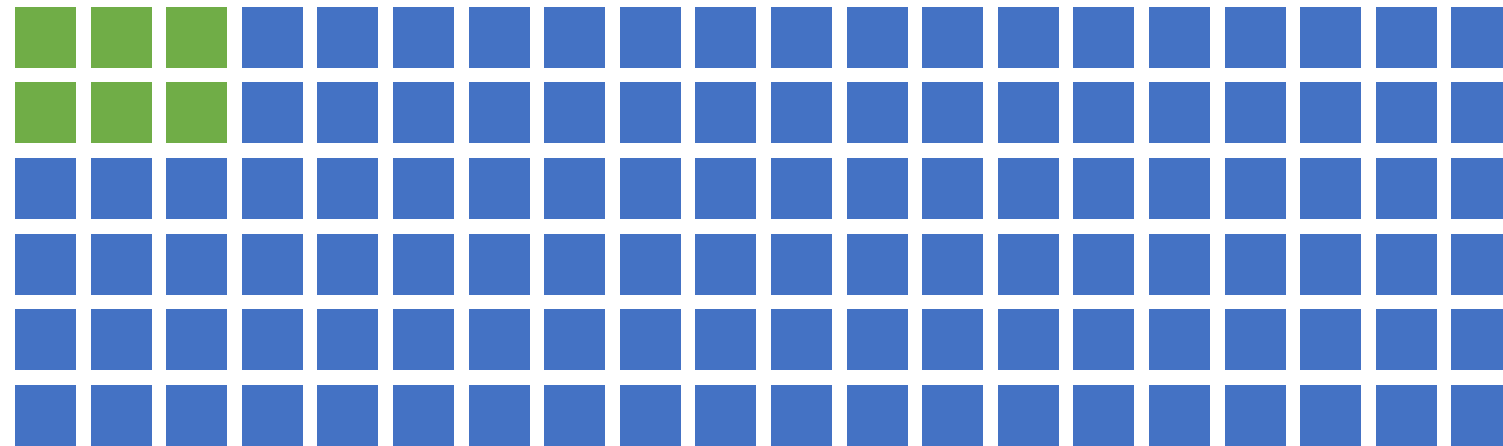
Para saber mais...



<https://www.casadocodigo.com.br/products/livro-big-data>

Considerações finais

Big Data está apenas em seu início



Estima-se que somente **0.5%**
dos dados globais são analisados

Considerações finais

BIG DATA =

- BIG INOVAÇÃO**
- BIG AGILIDADE**
- BIG CONHECIMENTO**
- BIG OPORTUNIDADE**

Um **BIG** Obrigada!

Meus dados 😊

Email: rpereira@larc.usp.br

LinkedIn: <https://br.linkedin.com/in/rosangelafpm>

Twitter: [hadoop_girl](#)

Lattes: <http://lattes.cnpq.br/1095618344499722>

Github: [rosangelapereira](#)



Sobre a autora



Rosângela de Fátima Pereira Marquesone é pesquisadora no Laboratório de Arquitetura e Redes de Computadores (LARC), atuando nas áreas de computação em nuvem e Big Data. Atua como professora em cursos de Big Data para empresas e programas de MBA e como revisora de código no nanodegree de Analista de Dados da rede de cursos on-line Udacity.

É autora do livro "Big Data - Técnicas e tecnologias para extração de valor de dados", publicado pela editora Casa do Código em 2016. É mestre e doutoranda em Engenharia de Computação pela Escola Politécnica da Universidade de São Paulo (Poli-USP).

Seus principais interesses de pesquisa são: Big Data, computação em nuvem, mineração de dados e Internet das coisas.

Também se interessa por temas como design thinking, mulheres na tecnologia e empreendedorismo social.