



Big Data

UMA INTRODUÇÃO

Introdução



O termo “big data” engloba dentro de si muitas definições



A ideia básica por trás dessa palavra é que tudo o que fazemos hoje está deixando uma pista digital (ou dado) o qual nós (e as outras pessoas) podem utilizar e analisar!



Desde o início da civilização até 2003, a humanidade gerou cinco **exabytes** de dados. Produzimos **cinco exabytes a cada dois dias...**

E o ritmo está acelerando.

(Eric Schmidt – Presidente Google)

Dados de Atividade

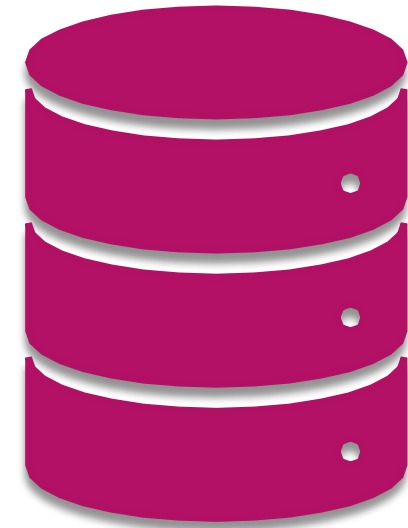
Qualquer atividade hoje "gera" dados!

Players de música digital e eBooks coletam dados das suas preferências.

Seu smartphone coleta dados sobre você ao usá-lo e seu navegador coleta informações sobre o que você está procurando.

Sua empresa de cartão de crédito coleta dados sobre **onde você compra** e a loja coleta dados sobre **o que você compra**.

É difícil imaginar qualquer atividade que não gere dados!



O que é Big Data?

Big data – grandes dados em Português – são similares aos dados normais (pequenos), mas em grande quantidade!

Por ser “grande” exige diferentes abordagens

- Técnicas, ferramentas e infraestrutura

O objetivo é o de resolver novos problemas ou problemas antigos de forma melhor

Big Data **gera valor** a partir do armazenamento e processamento de quantidades muito grandes de informações digitais que não podem ser analisados com técnicas tradicionais de computação.



Dataficação!

BigData Analytics
DataWarehouse PowerBI
DataQuality BLOG DATAFICAÇÃO
DataMining BI
Microsoft OLAP

- ▶ Como consequência do big data surge a **dataficação**
- ▶ **Dataficação** é a tendência tecnológica moderna de transformar diversos aspectos de nossa vida em dados que são posteriormente transformados em informação percebida como **uma nova forma de valor**
<https://pt.wikipedia.org/wiki/Dataficação>
- ▶ Big data, que é muitas vezes descrito usando quatro V's

Volume

Quantidade de dados

Velocidade

Tráfego de dados

Variedade

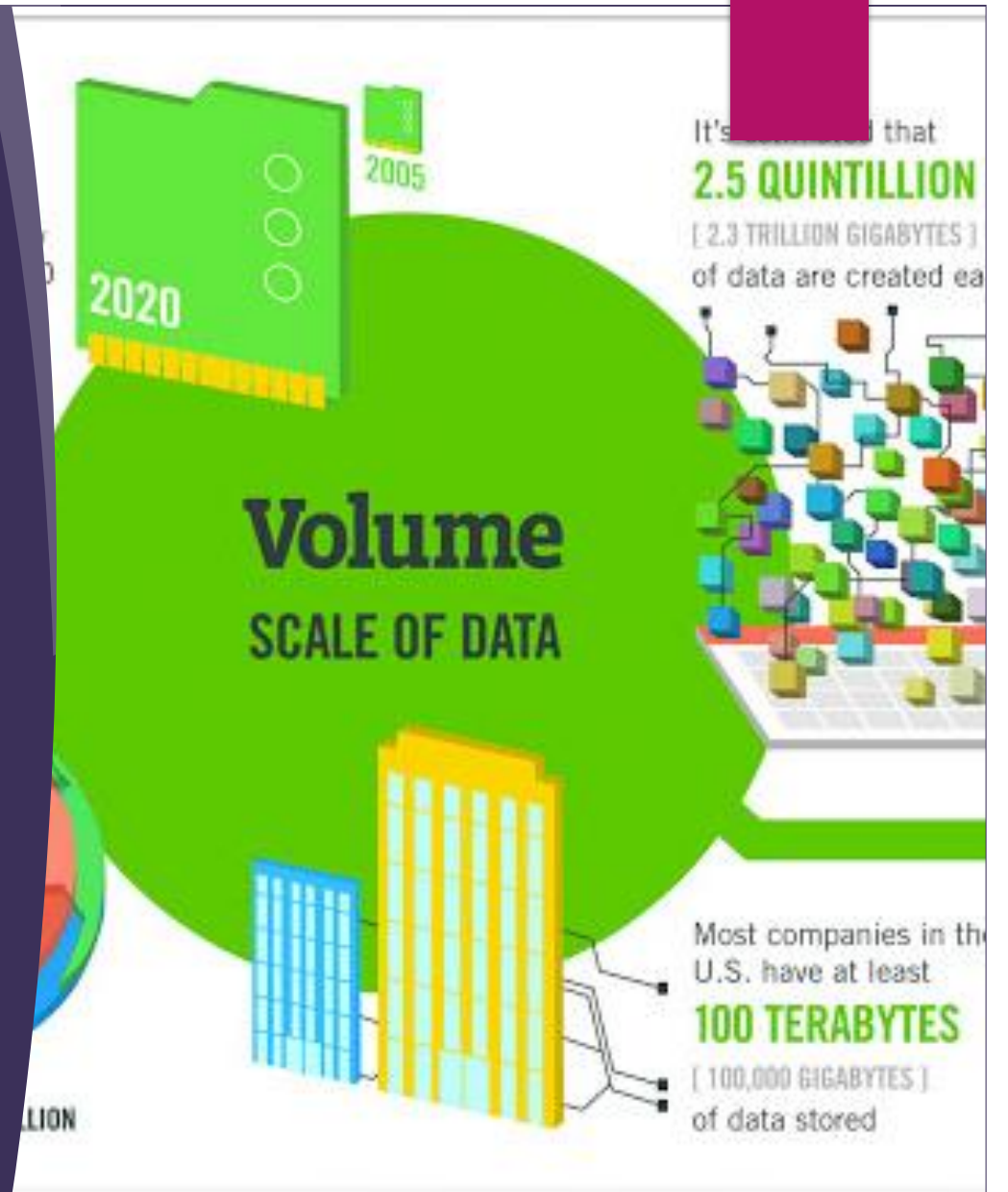
Tipos de dados

Veracidade

Confiabilidade dos dados

Volume

- ▶ Refere-se às vastas quantidades de dados gerados a cada segundo
- ▶ Não se fala mais em Terabytes, mas Zettabytes ou Brontobytes
- ▶ Se tomarmos todos os dados gerados no mundo entre o início dos tempos e os anos 2000, a mesma quantidade de dados serão gerados em breve a cada minuto
- ▶ Alguns exemplos:
 - ▶ Um PC típico tinha 10 gigabytes de armazenamento em 2000
 - ▶ Hoje, o Facebook consome 500 terabytes de novos dados todos os dias
 - ▶ Um Boeing 737 gera 240 terabytes de dados durante um único voo através dos EUA
 - ▶ Sensores embutidos em objetos cotidianos em breve irão gerar bilhões de novos feeds constantemente atualizados sobre o meio ambiente, contendo dados de localização e outras informações, incluindo vídeo.
- ▶ Novas ferramentas de big data usam sistemas distribuídos para armazenar e analisar dados em bancos de dados espalhados em qualquer lugar em o mundo.



Velocidade

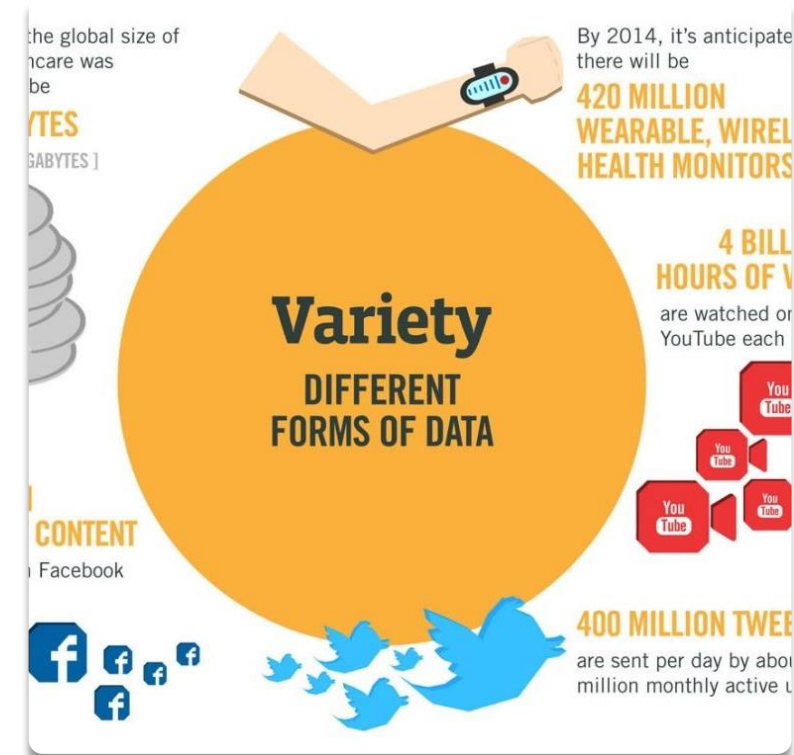
- ▶ ... refere-se à velocidade em que novos dados são gerados e se deslocam.
- ▶ Basta pensar em mensagens de mídia social se tornando virais em segundos.
- ▶ A tecnologia permite analisar os dados enquanto estão sendo gerados (às vezes chamados de análises em memória), sem nunca colocá-los em bancos de dados.
- ▶ Alguns exemplos:
 - ▶ Sequências de cliques e taxa de **cliques** de propaganda capturam o comportamento dos usuários em milhões de eventos por segundo
 - ▶ algoritmos de negociação de ações de alta frequência refletem mudanças no mercado em microssegundos
 - ▶ Redes de sensores geram maciços registros de eventos relevantes (logs) em tempo real
 - ▶ Sistemas de jogos on-line comportam milhões de usuários simultâneos, cada um produzindo várias entradas por segundo.



Internet
of Things
& Big Data

Variedade

- ▶ ... refere-se aos diferentes tipos de dados que agora podemos usar
- ▶ Big Data não são apenas números, datas e caracteres
- ▶ Também são dados geo espaciais, dados 3D, áudio e vídeo, e texto não estruturado, incluindo arquivos de log e redes sociais.
- ▶ 80% dos dados do mundo são não estruturados (texto, imagens, vídeo, voz, etc.)
- ▶ No passado, focava-se apenas em dados estruturados que se encaixavam perfeitamente em tabelas ou bancos de dados relacionais, como dados financeiros
- ▶ Os sistemas tradicionais de banco de dados foram projetados para atender a volumes menores de dados estruturados, poucas atualizações ou uma estrutura de dados previsível e consistente.
- ▶ Com a tecnologia de big data, agora podemos analisar e reunir dados de diferentes tipos, como mensagens, conversas em redes sociais, fotos, dados de sensores, vídeos ou gravações de voz



Veracidade



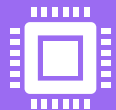
... refere-se à confiabilidade dos dados.



As formas, a qualidade e a precisão no big data são menos controláveis



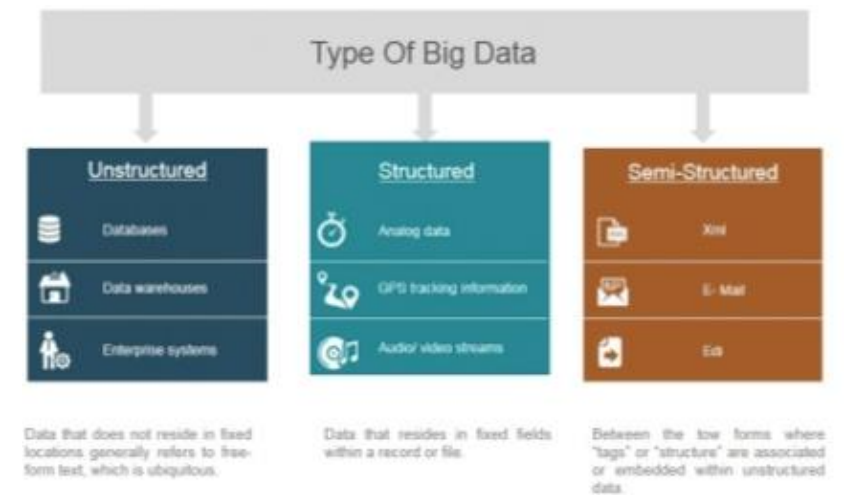
Basta pensar em Posts no Twitter com hash tags, abreviaturas, erros de digitação e discurso coloquial, bem como a confiabilidade e a precisão dos conteúdos ...



Porém, a tecnologia atual permite trabalhar com este tipo de dados.

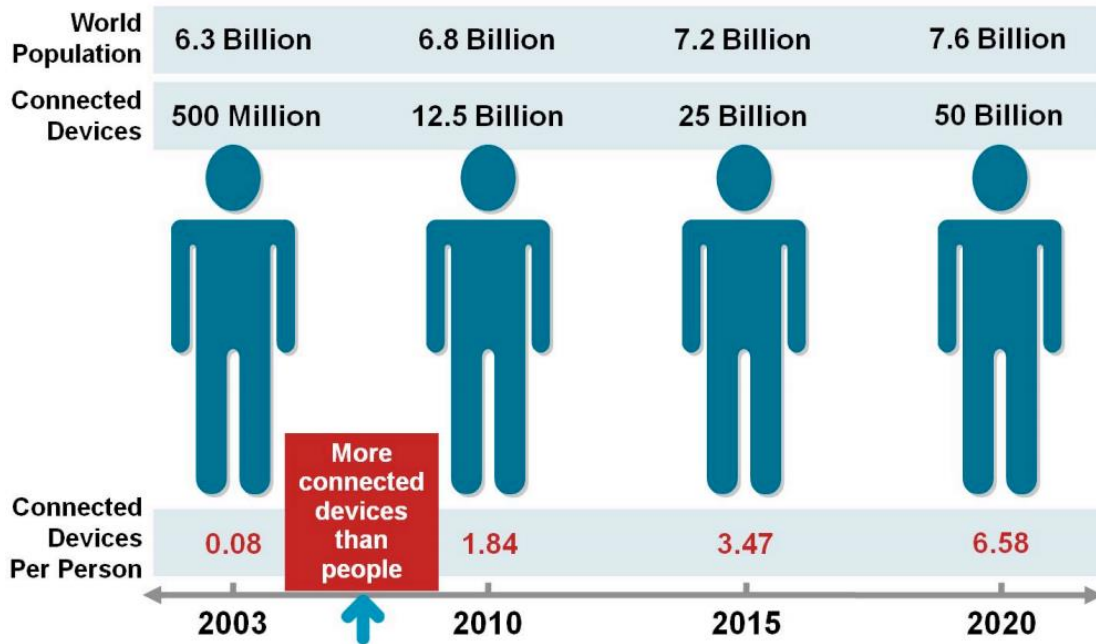
Tipos de Dados

- ▶ Estruturados
 - ▶ A maioria das fontes de dados tradicionais
- ▶ Semiestruturados
 - ▶ Muitas das fontes de dados do big data
- ▶ Não estruturados
 - ▶ Texto, imagens, vídeo, voz e etc.

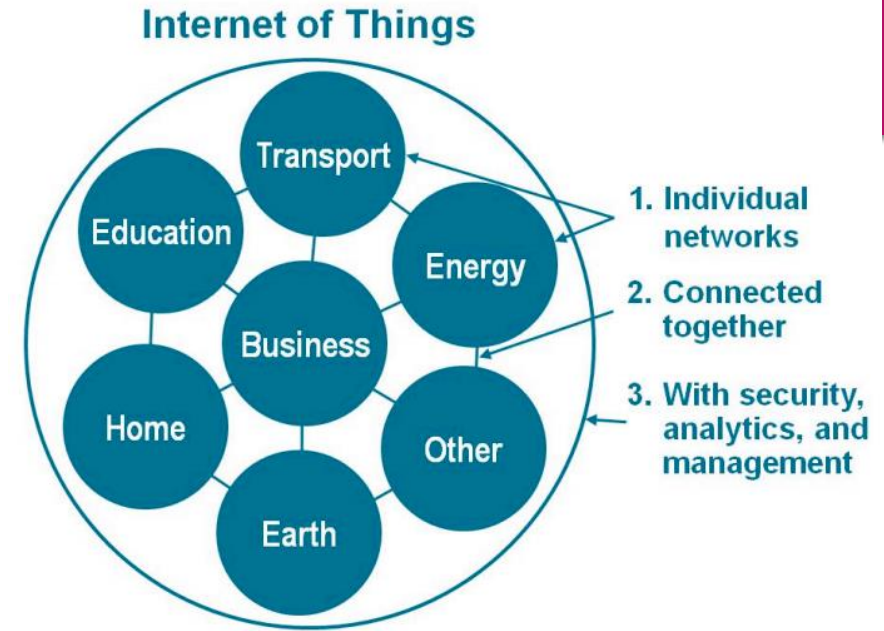


Por que Big data ?

- ▶ O big data já é uma realidade e uma necessidade!
- ▶ Conjunturas favoráveis:
 - ▶ Aumento da capacidade de armazenamento
 - ▶ Aumento do poder de processamento
 - ▶ Disponibilidade de diferentes tipos de dados
 - ▶ Todos os dias são criados 2,5 quintilhões de bytes de dados
- ▶ 90% dos dados no mundo hoje foram criados apenas nos últimos dois anos



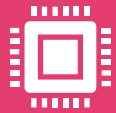
Source: Cisco IBSG, April 2011



Source: Cisco IBSG, April 2011

Por que Big Data ?

Em que o Big Data é diferente?



Dados gerados automaticamente por uma máquina (exemplo: sensor embutido em um motor)



Normalmente uma fonte de dados totalmente nova (exemplo: uso da internet)



Não foi projetado para ser amigável (exemplo: fluxos de texto (text stream))



Pode não ter muitos valores pois concentra-se na parte importante

Fontes de Big Data



Usuários



Aplicativos

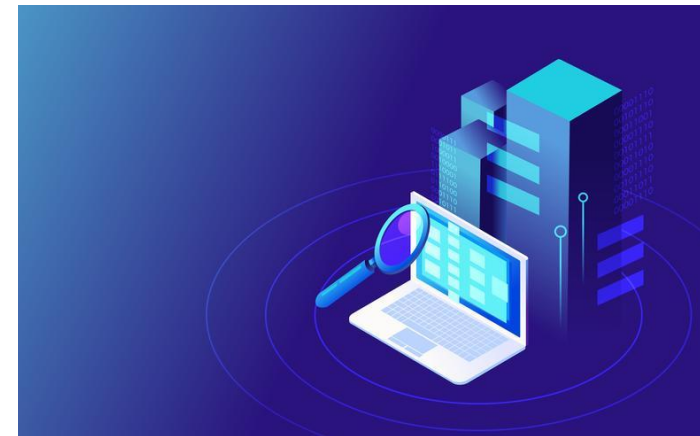


Sistemas



Sensores

Arquivos grandes e em constante crescimento
(Arquivos de big data)



Big Data Analytics

Definição

Análise de big data é o processo complexo de examinar conjuntos de dados grandes e variados, ou big data, para descobrir informações, tais como padrões ocultos, correlações desconhecidas, tendências de mercado e preferências do cliente, que podem ajudar as organizações a tomar decisões de negócios inteligentes.

(<https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>)

Objetivos

- ▶ Examinar grande quantidade de dados
- ▶ Informações adequadas
- ▶ Identificar padrões ocultos, correlações desconhecidas
- ▶ Criar vantagem competitiva
- ▶ Obter melhores decisões de negócios: estratégicas e operacionais
- ▶ Oferecer marketing eficaz, satisfação do cliente, aumento da receita



Aplicações de Big Data Analytics

**Assistência
Médica
Inteligente**



**Vendas por
multicanais**



**Cidades
Inteligentes**



**Segurança
Nacional**



Telecomunicações



**Análise da
Aprendizagem**



**Controle de
Tráfego**



**Análise de
Negociações**



**Sistema de
Recomendação**



**Indústria
manufatureira**



**Qualidade de
Busca**



Games



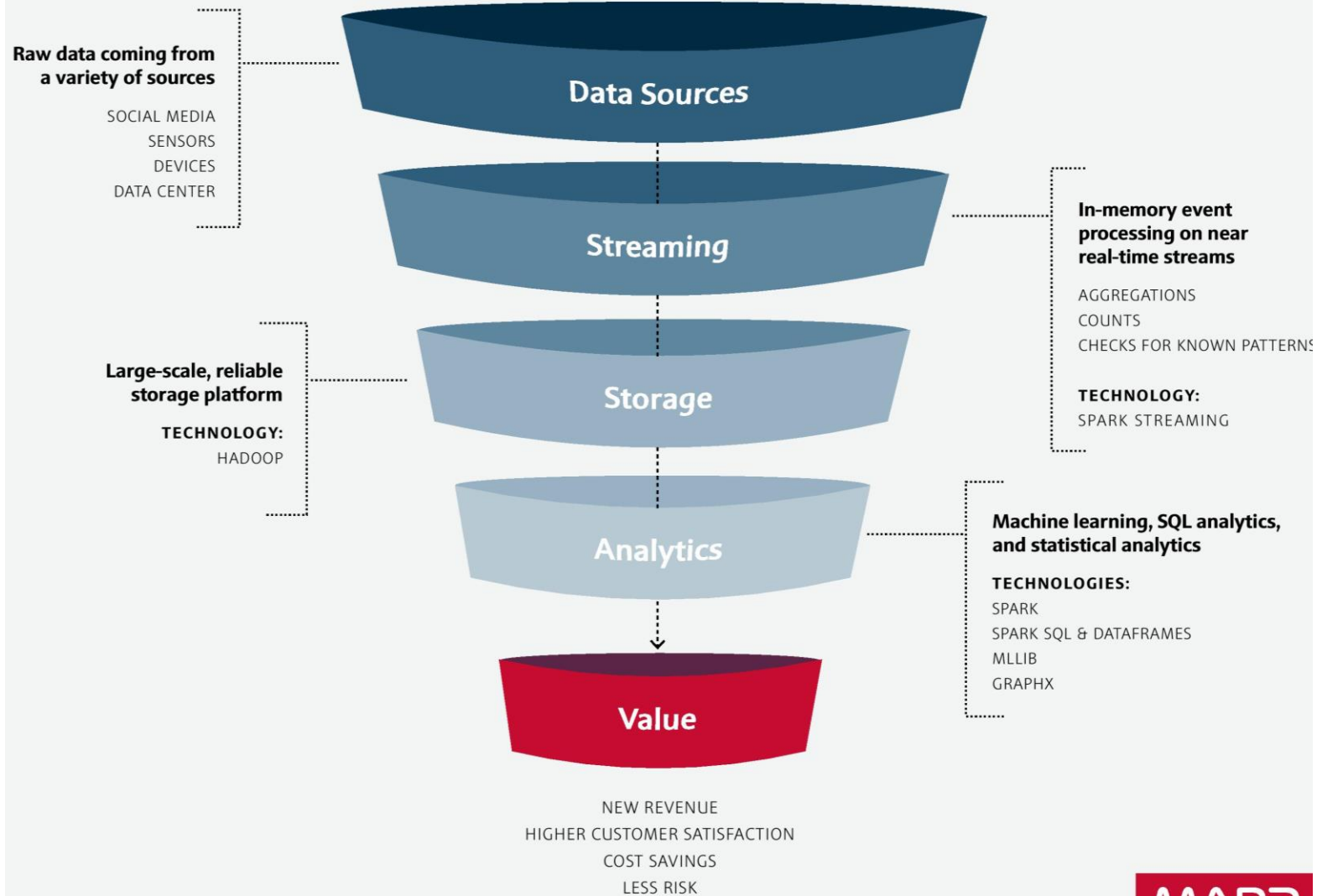
Transformando o Big Data em Valor

- ▶ A **dataficação** do mundo oferece quantidades sem precedentes de dados em termos de Volume, Velocidade, Variedade e Veracidade.
- ▶ As mais recentes tecnologias, tais como computação em nuvem e sistemas distribuídos juntamente com as mais recentes abordagens em software de análise permitem alavancar todos os tipos de dados para obter insights e agregar valor.

Transformando o Big Data em Valor



Turning Data into Value with Spark and Hadoop



Transformando
o Big Data em
Valor

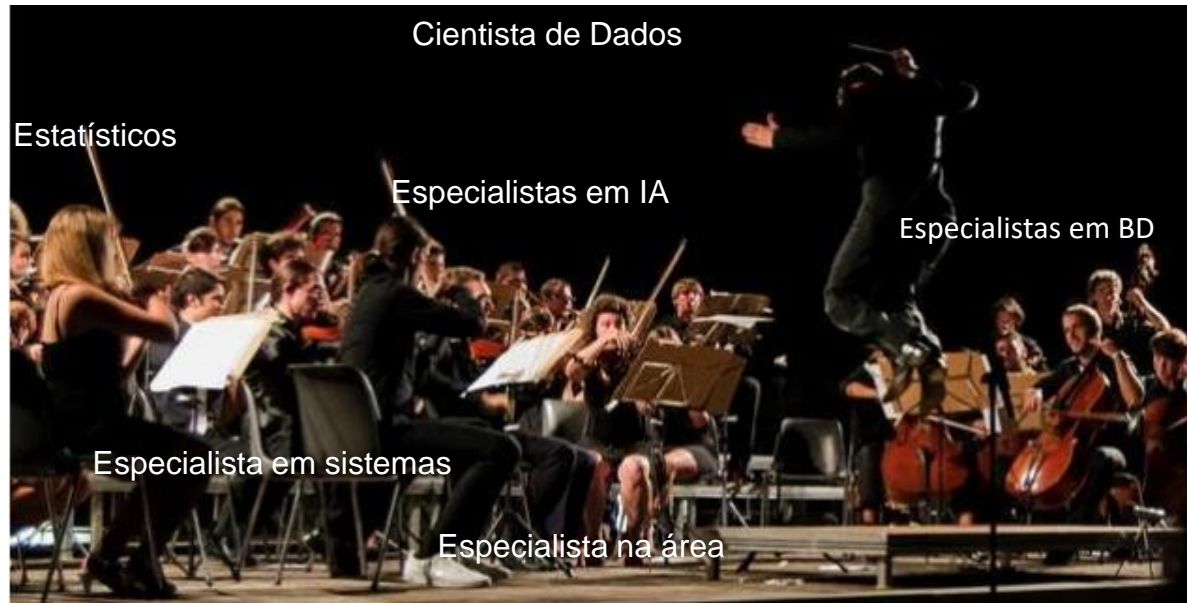
Mas as aplicações em big data são infinitas ...

- ▶ Atualmente estamos apenas vendo o início de uma transformação em uma economia de big data
- ▶ Qualquer negócio que não considere seriamente as implicações do Big Data corre o risco de ser deixado para trás.

Bernard Marr. Big Data in Practice, Wiley, 2016.

Ciclo de Vida de Big Data

- ▶ Executar o ciclo de vida de big data é difícil!
- ▶ Precisa de amplo entendimento em muitos campos
- ▶ Equipes de big data incluem membros de diferentes áreas trabalhando em conjunto!



Dark Side

► Privacidade

- Invasão de privacidade
- Dados podem ser utilizados em coisas inesperadas

► Big Brother

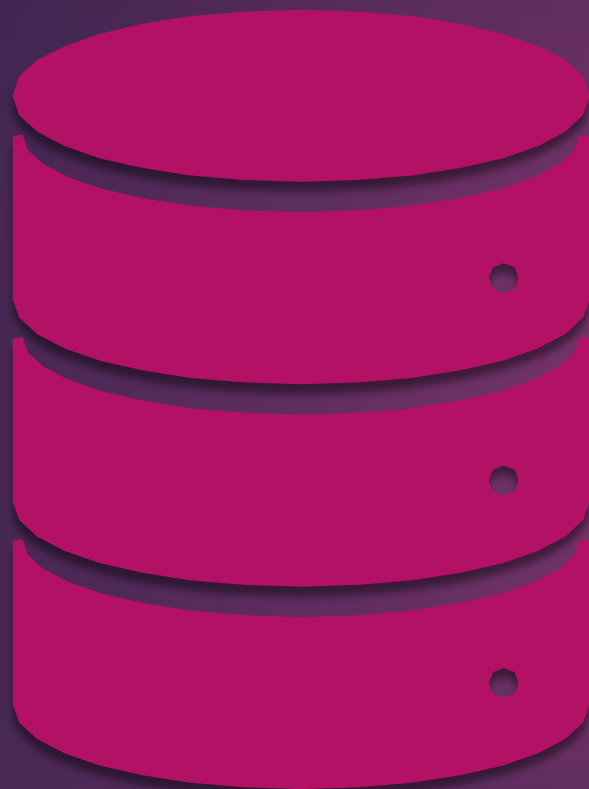
- Dados utilizados provavelmente para controle (p.ex. governos!)
- Se a tecnologia está lá fora, pode ser que esteja tudo bem. É muito difícil esconder alguma coisa, isto funciona em ambos os sentidos...



Concluindo ...

- ▶ Existem muitos dados disponíveis, e dados ➡ insights ➡ previsões
- ▶ São realizadas muitas previsões atualmente
- ▶ Existe muuita relação entre os dados e as previsões
 - ▶ Análise
 - ▶ Visualizações
 - ▶ Padrões
 - ▶ Mineração de Dados
- ▶ Visualização é o ponto-chave!!





Big data

UMA INTRODUÇÃO