
BigData

Fundamentos e Desafios

Profa. Andreza Leite

BigData

- Pra que serve e quem usa?
- É diferente de BI?
- Porque esta onda toda agora?
- Como funciona na prática?
- Quais os desafios?

Big Data: enxurrada de dados emerge como novo termômetro da economia

- Empresas e governos exploram a tecnologia para analisar em tempo real indicadores como inflação e emprego, uma metodologia que ainda enfrenta obstáculos

 Recomendar

15

 Tweet



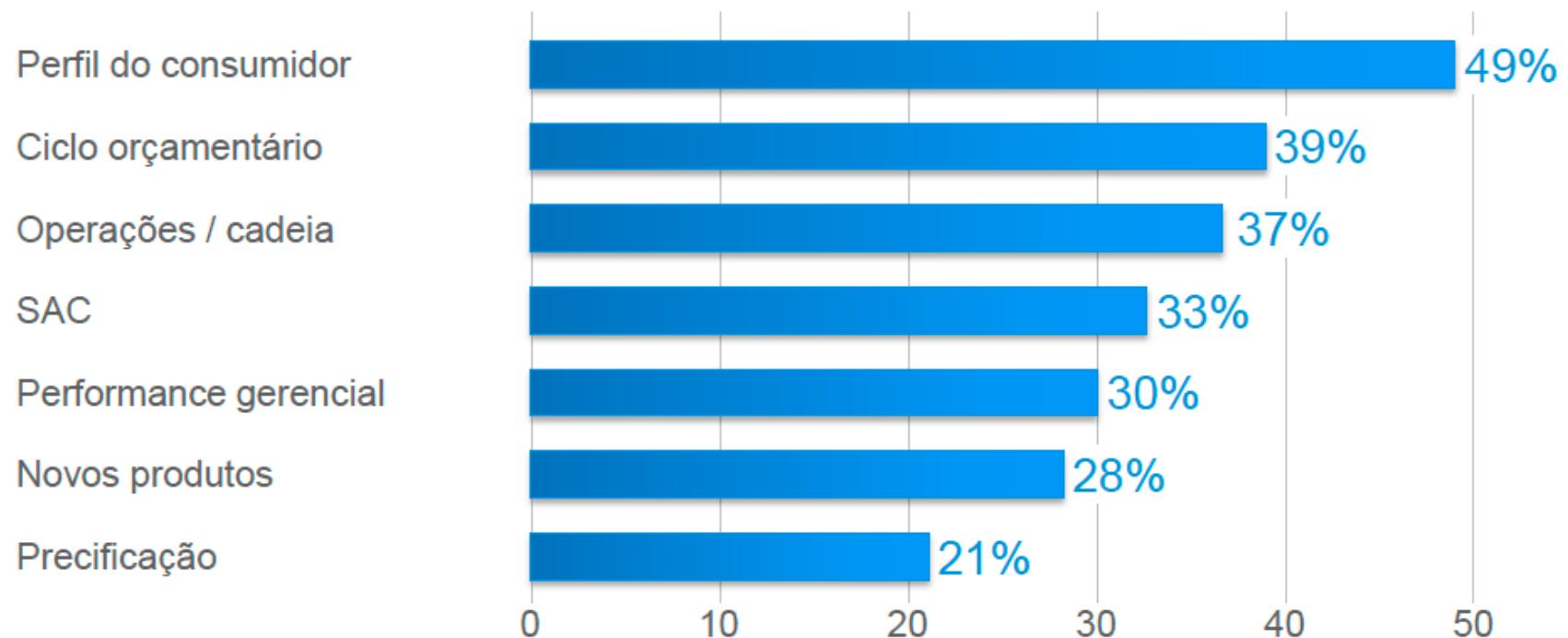
RENNAN SETTI (EMAIL)

Publicado: 22/03/14 - 21h45 Atualizado: 23/03/14 - 13h27

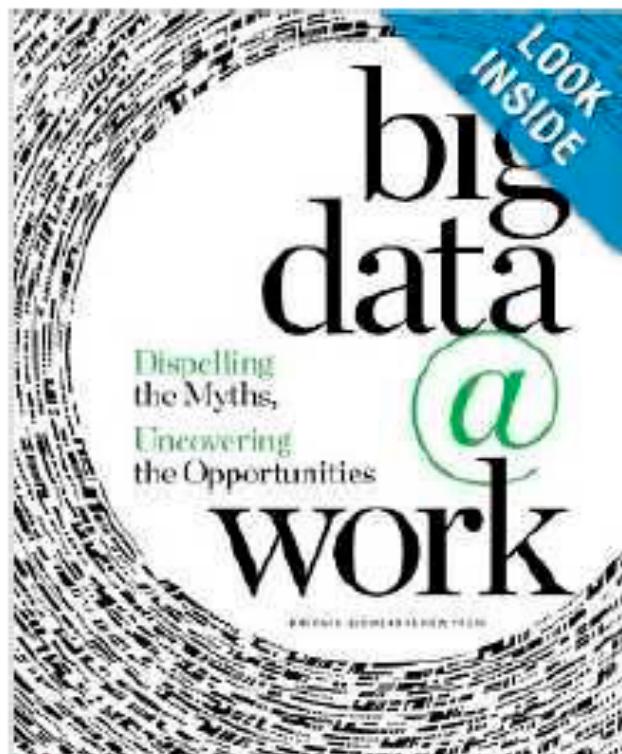


Utilização

Áreas funcionais onde estão utilizando BigData



McKinsey Global Survey of 1,469 C-level executive respondents at a range of industries and company sizes, "Minding Your Digital Business," 2012.



THOMAS H. DAVENPORT
New York Times bestselling author of *Competing on Analytics*

HBIG.ORG

Harvard Business Review

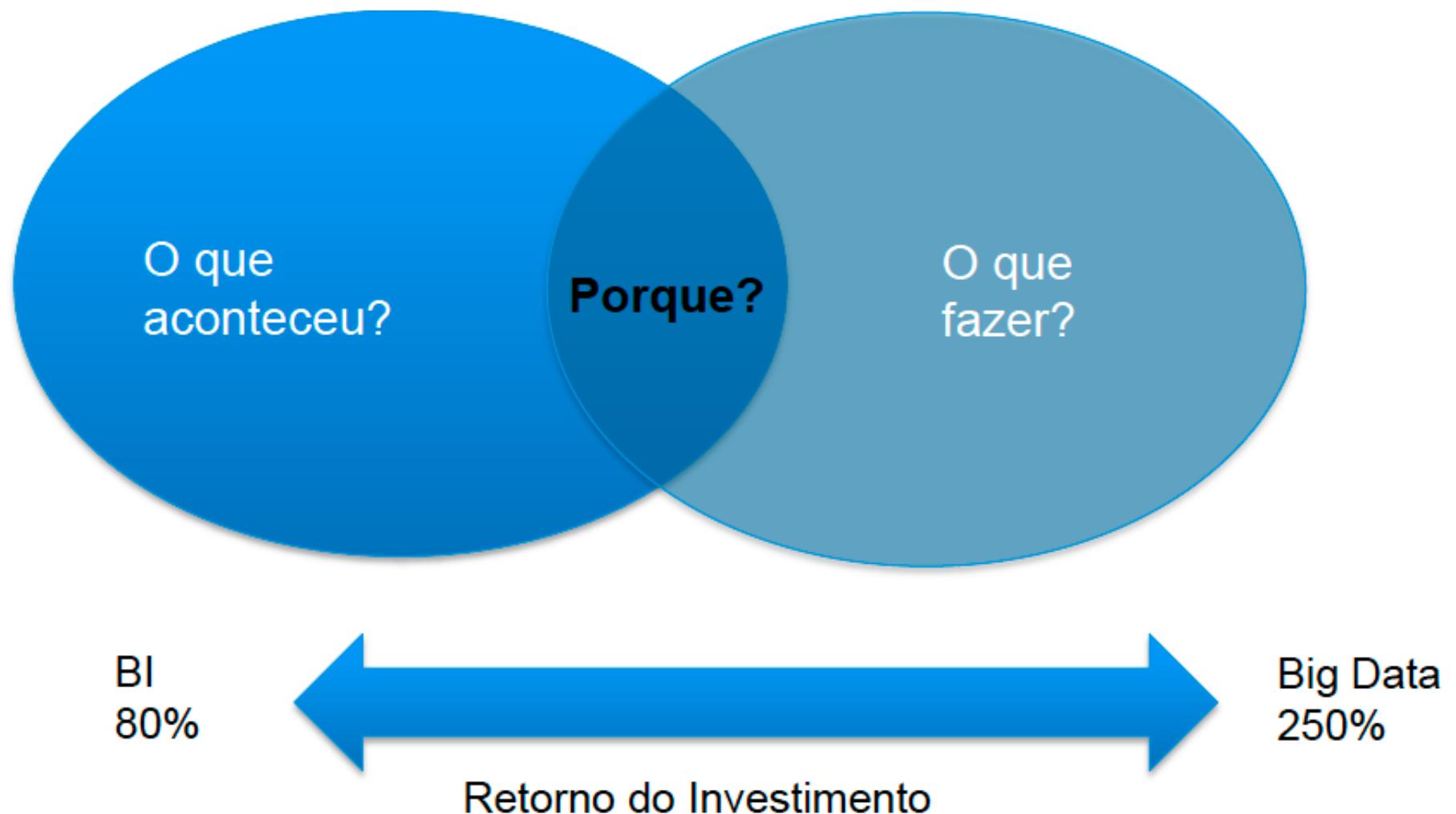
OCTOBER 2012
REPRINT R12102

SPOTLIGHT ON BIG DATA

Data Scientist: The Sexiest Job Of the 21st Century

Meet the people who can coax treasure
out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

BI x BigData



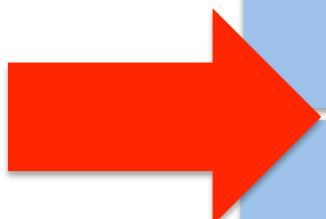
BI

BANCO RELACIONAL

ETL

DADOS BRUTOS

COLETA DE DADOS



BI x BigData



80%

250%

MUNDO DIGITAL



2000: 2 Exabytes por ano
2011: 2 Exabytes por dia

Em 2020, a Internet vai conectar
7.6B de pessoas e **200B** de objetos
(sensores, máquinas, dispositivos....)

Menos de **1%** dos dados são
analizados

E no Brasil?

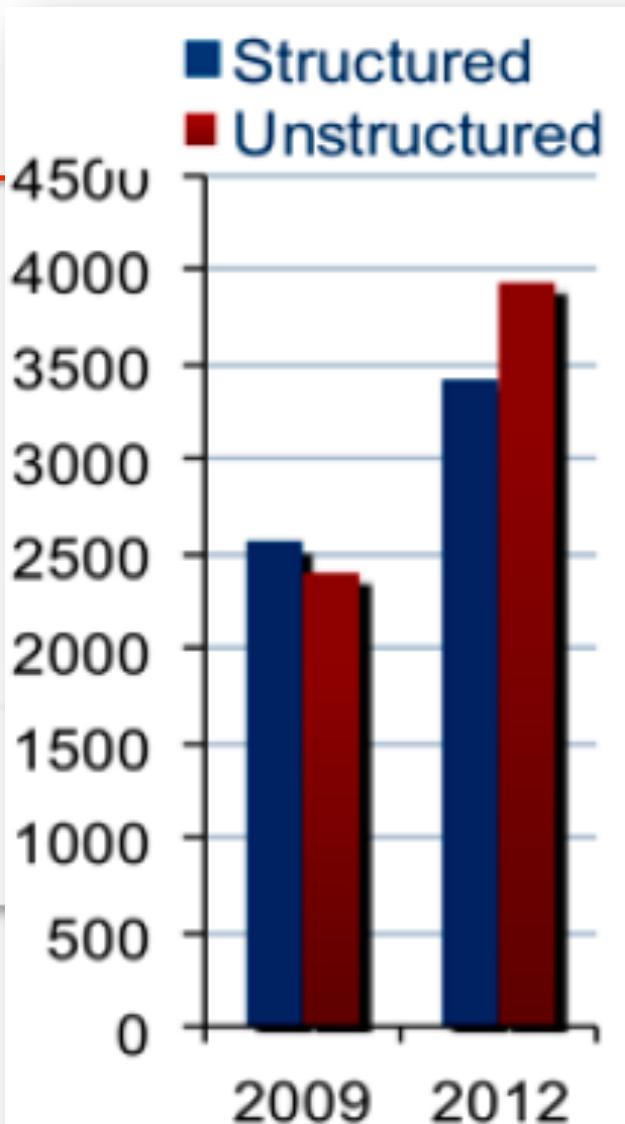


<http://brazil.emc.com/leadership/digital-universe/index.htm>

-
- ↑Aumento continuado do consumo de smartphones, internet e redes sociais;
 - ↑Investimento agressivo de companhias em TI para aumentar sua competitividade;
 - ↓Queda nos custos de tecnologias que capturam, gerenciam, protegem e armazenam informações; e
 - ↑Crescimento da comunicação machine2machine e as informações sobre informações.

Antes ...

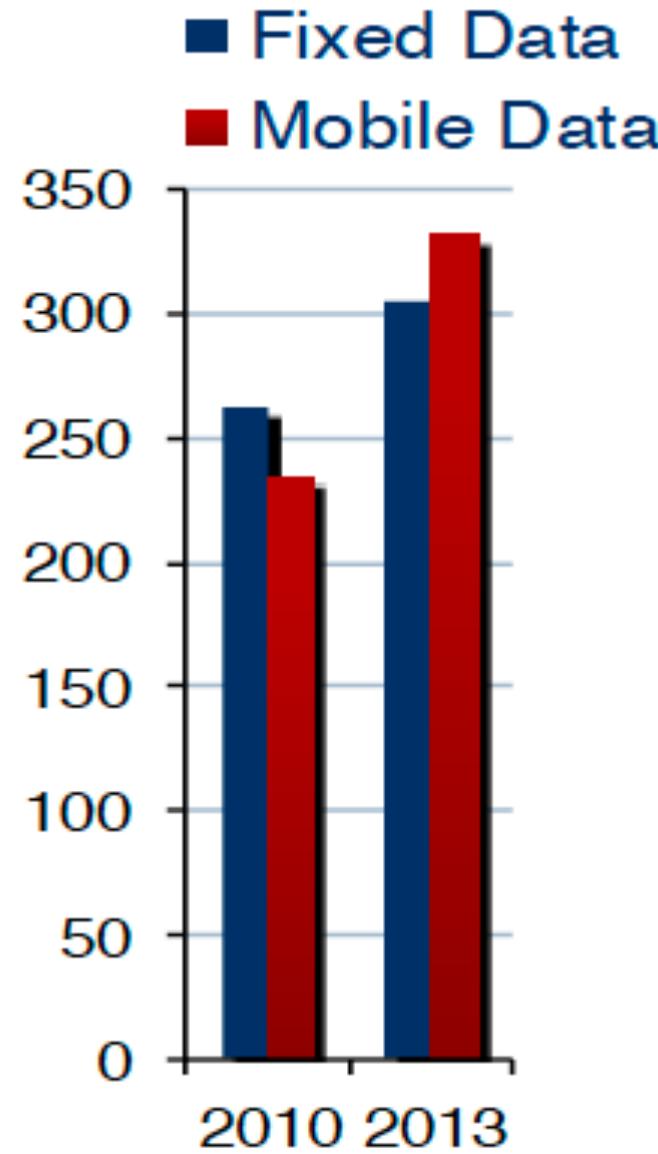
```
1 CREATE DATABASE Consultorio
2 USE Consultorio
3
4 CREATE TABLE Medico
5 (
6     IDMEDICO      INT IDENTITY PRIMARY KEY NOT NULL,
7     NOMEMEDICO   VARCHAR(50),
8     CRM          VARCHAR(10)
9 )
10
11 CREATE TABLE Paciente
12 (
13     IDPACIENTE    INT IDENTITY PRIMARY KEY NOT NULL,
14     NOMEpaciente  VARCHAR(50),
15     TELEFONE      VARCHAR(10)
16 )
17
18 CREATE TABLE Consulta
19 (
20     IDCONSULTA    INT IDENTITY PRIMARY KEY NOT NULL,
21     IDMEDICO      INT NOT NULL,
22     IDPACIENTE    INT NOT NULL,
23     DATACONSULTA  DATETIME,
24     HORAINICIO    DATETIME,
25     HORAFIM       DATETIME,
26     OBSERVACOES   VARCHAR(MAX),
27     ATIVO         BIT
28 )
```



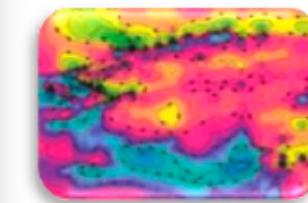
ail™
BETA

Photoshop®
See What's Possible®

Não-corporativo



Source: 2013 IDC Digital Universe Study







BigData



Variedade de tipos

- Estruturados
 - Dados em um SGBD
 - Esquema
- Não-estruturados
 - Tweets, logs, vídeos, som, imagens
- Semi-estruturados
 - XML, JSON, RDF, OWL



Estruturados

- ❑ Dados organizados em blocos semânticos (relações)
- ❑ Dados de um mesmo grupo possuem as mesmas descrições (atributos)
- ❑ Descrições para todas as classes de um grupo possuem o mesmo formato (esquema)
- ❑ Dados mantidos em um SGBD são chamados de Dados Estruturados por manterem a mesma estrutura de representação (rígida), previamente projetada (esquema)

Semi-Estruturados

- ✓ Dados Web são bastante heterogêneos
 - ✓ A alta heterogeneidade dificulta as consultas a estes dados
-
- ❖ Dados semi-estruturados
 - ❖ Não são estritamente tipados
 - ❖ Não são complementamente não-estruturados
-
- ❖ O esquema de representação está presente (de forma explícita ou implícita)
-
- ❖ Auto-descritivo
-
- ❖ Uma análise do dado deve ser feita para que a sua estrutura possa ser identificada e extraída

Semi-Estruturados

(Características principais)

- ❖ Definição à posteriori
 - ❖ Esquemas são definidos após a existência dos dados
 - ❖ Investigação de suas estruturas particulares
- ❖ Estrutura irregular
 - ❖ Não existe um esquema padrão para os dados
 - ❖ Coleções de dados são definidos de maneiras diferentes, contendo informações incompletas
- ❖ Estrutura implícita
 - ❖ Muitas vezes existe uma estrutura implícita
- ❖ Estrutura parcial
 - ❖ Apenas parte dos dados disponíveis podem ter uma estrutura

Dados Estruturados	Dados Semi-Estruturados
Esquema pré-definido	Nem sempre há um esquema
Estrutura regular	Estrutura irregular
Estrutura independente dos dados	Estrutura embutida nos dados
Estrutura reduzida	Estrutura extensa (particularidades de cada dado, visto que cada um pode ter uma organização própria)
Fracamente evolutiva	Fortemente evolutiva (estrutura modifica-se com frequência)
Prescritiva (esquemas fechados e restrições de integridade)	Estrutura descritiva
Distinção entre estrutura e dados é clara	Distinção entre estrutura e dados não é clara

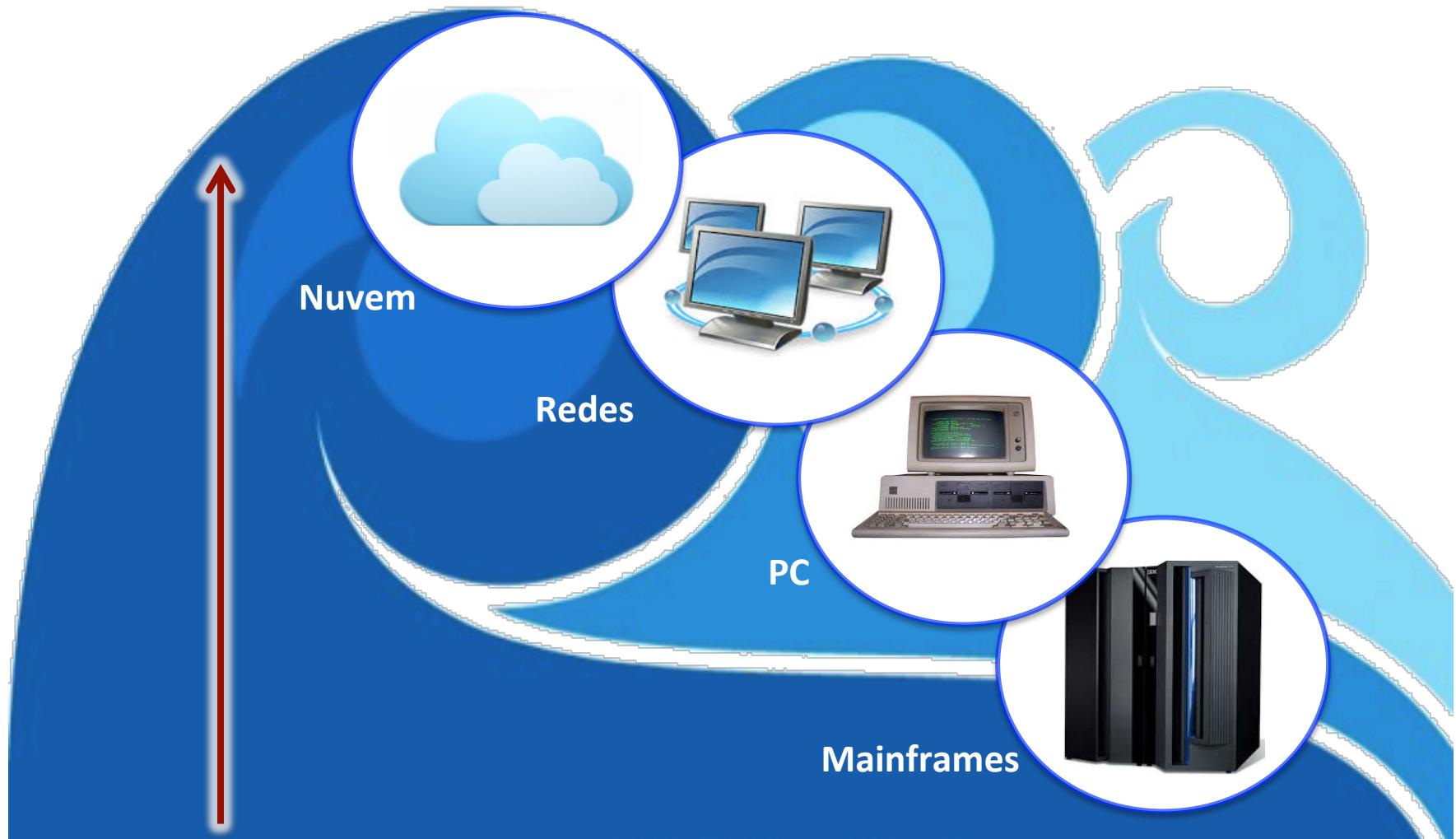
Não Estruturados

- ✧ São os dados que não possuem uma estrutura definida.
- ✧ Normalmente caracterizados por documentos textos, imagens, vídeos, etc
- ✧ Estruturas não são descritas implicitamente
- ✧ Grande maioria dos dados atuais na Web e nas empresas seguem este formato.

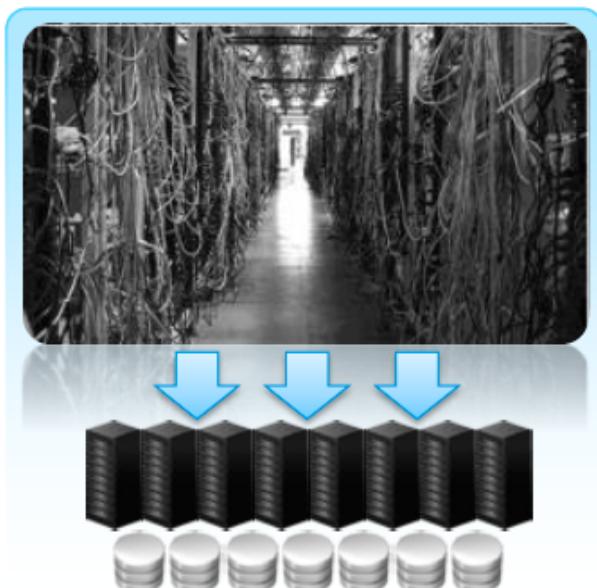
O que é diferente agora?



**Natureza
dos Dados**



Computação em Nuvem



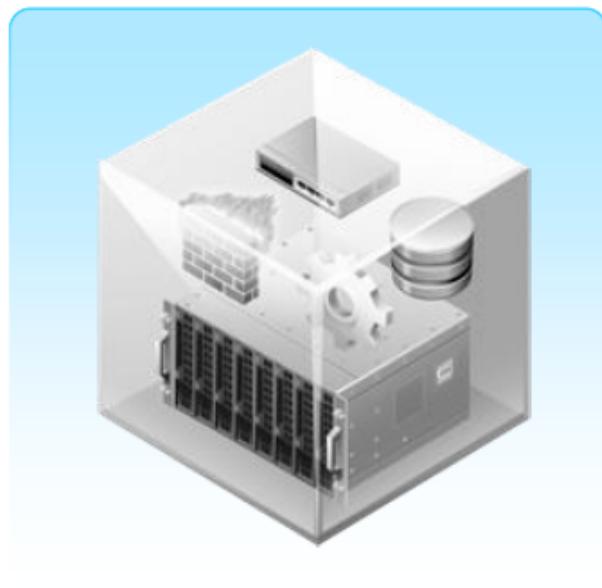
1

Padroniza



2

Virtualiza



3

Automatiza

O que é diferente agora?



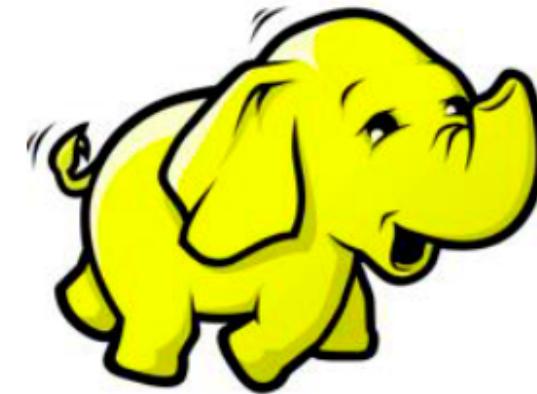
**Natureza
dos Dados**



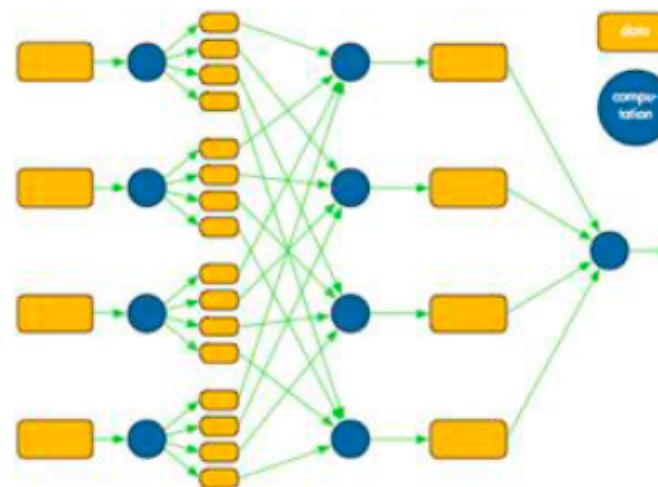
**Economia
de Escala**



EMC²



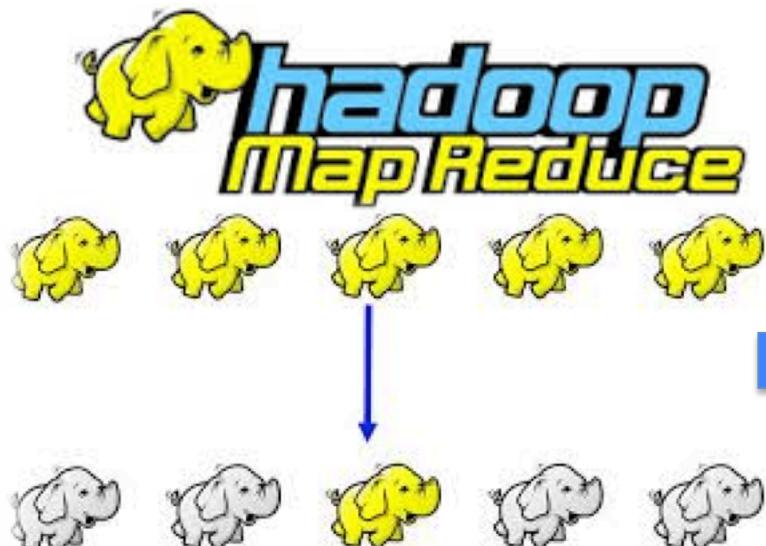
YAHOO!



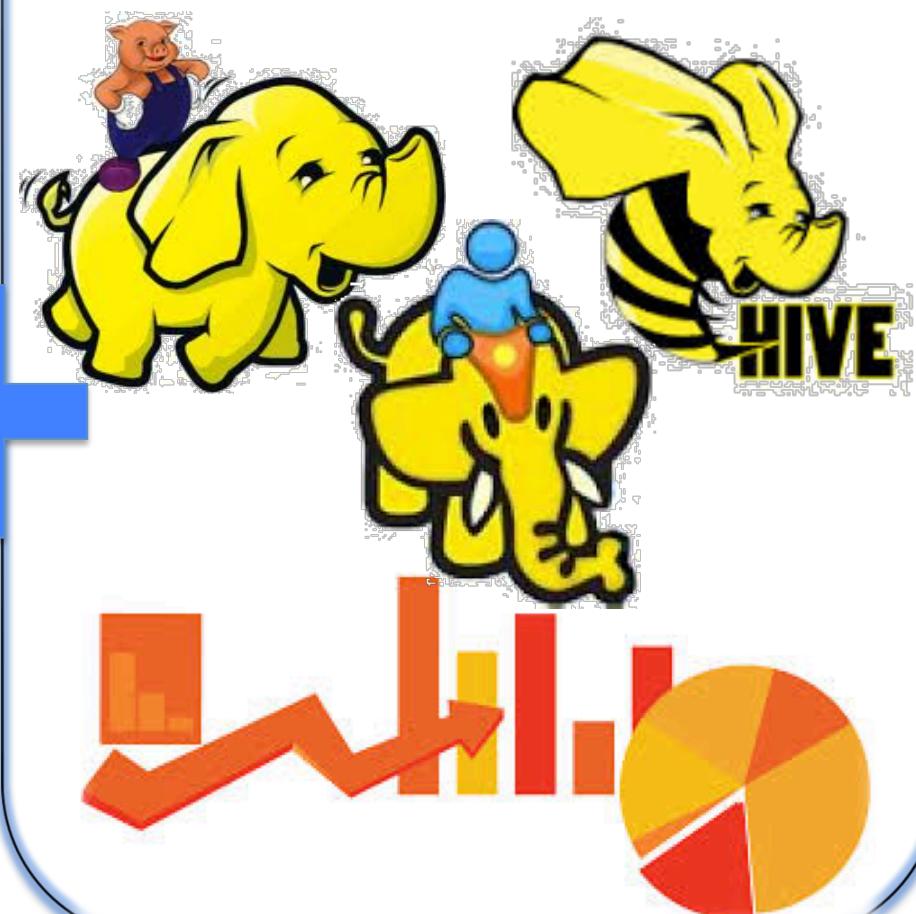
Map Reduce

Google

Framework



Interfaces de alto nível



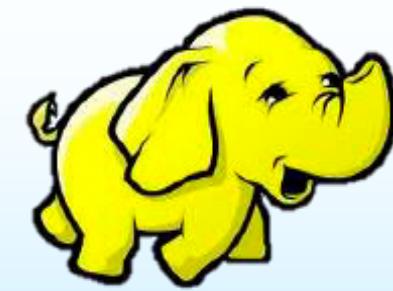
O que é diferente agora?



**Natureza
dos Dados**

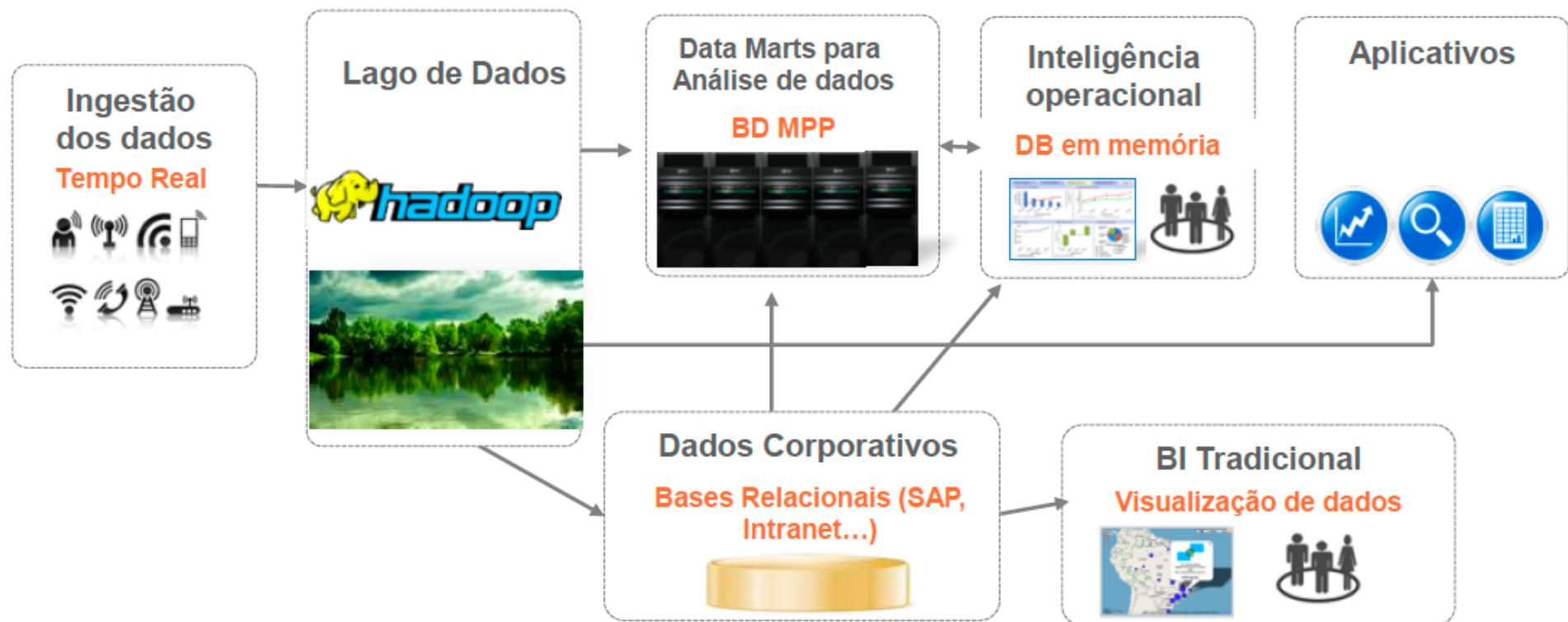


**Economia
de Escala**



**Novas
Tecnologias**

Arquiteturas



-
- Quem é o dono dos dados?
 - Quem disponibiliza a infraestrutura?
 - Qual o perfil da mão de obra para gerar/
manter aplicativos?
 - cientistas de dados

Data Scientist:

The Sexiest Job of the 21st Century

“...faltam cientistas de
dados”



INTERNATIONAL INSTITUTE FOR ANALYTICS.

2013 Predictions