



**Instituto Superior de Economia e Gestão**

---

UNIVERSIDADE TÉCNICA DE LISBOA

# Big Data

Rui Rosa

Aula ISEG  
14-Nov-2014

# Big Data

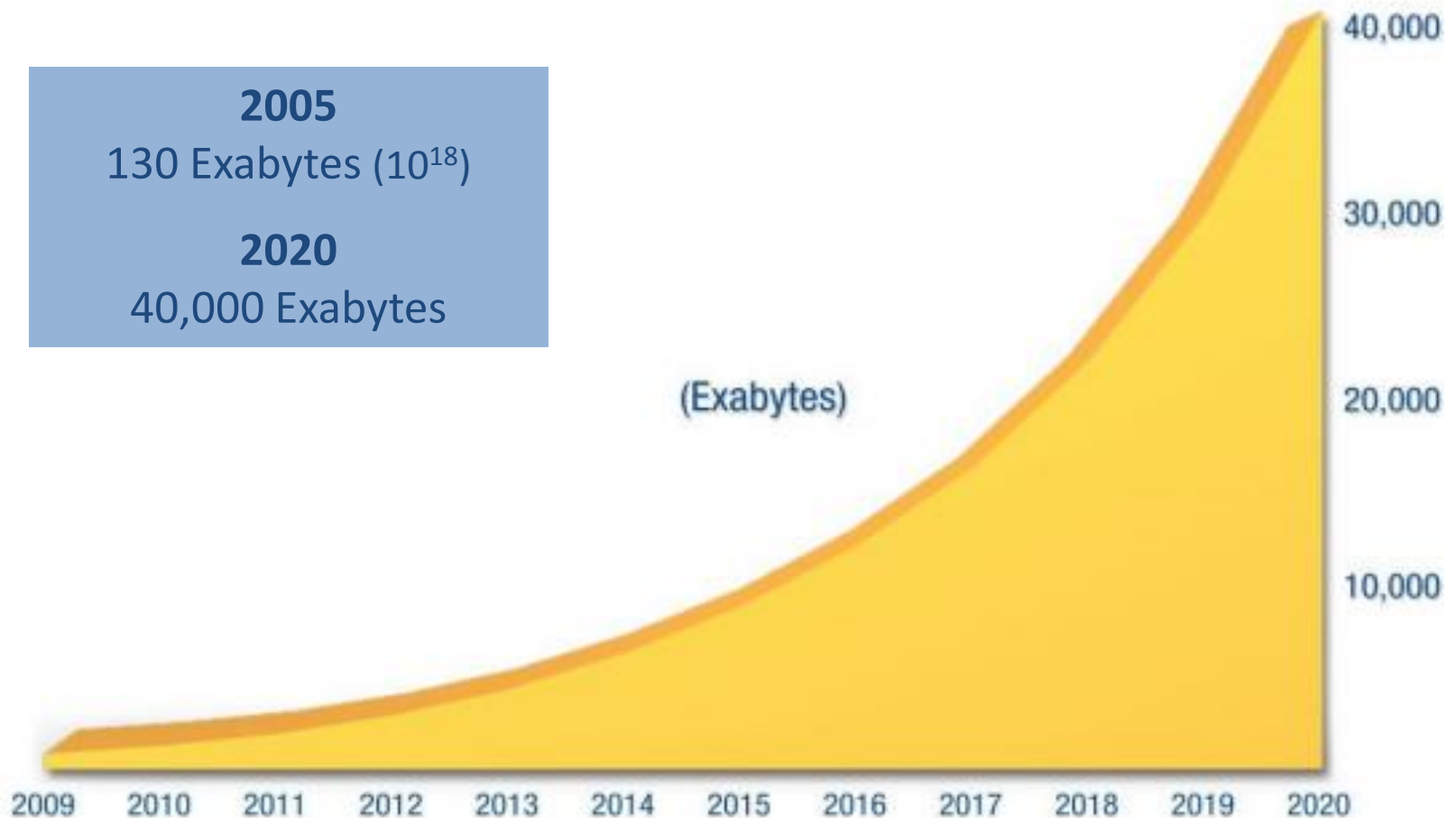
Uma história de inovação. A acontecer!

1. O que é “Big Data” e porque é importante
2. Tecnologias e arquitetura
3. Processos e pessoas
4. Adoção e vantagens
5. Inovação, segurança e ética
6. Conclusões

**O QUE É E PORQUE É IMPORTANTE**

# Big Data

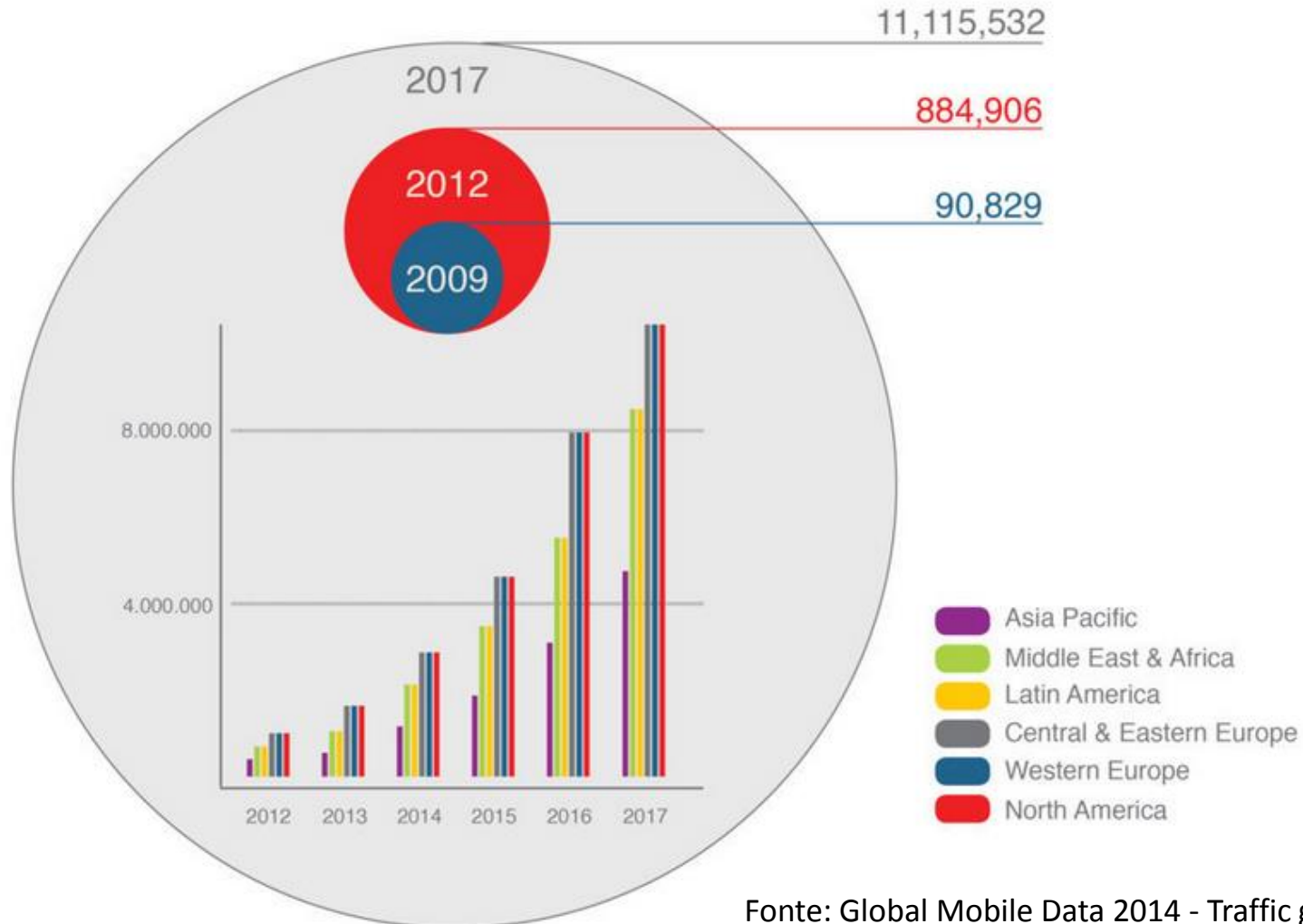
The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

# Big Data

Global Mobile Data - Traffic growth & forecast (terabytes per month)



Fonte: Global Mobile Data 2014 - Traffic growth and forecast

# Big Data

## Informação

- “Every day, we create 2.5 quintillion (Ex,  $10^{18}$ ) bytes of data”  
IBM (2013); <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- “We create as much information in two days now as we did from the dawn of man through 2003.”  
Eric Schmidt at Techonomy (2010); <http://techcrunch.com/2010/08/04/schmidt-data/>
- “A full 90% of all the data in the world has been generated over the last two years.”  
SINTEF (2013); <http://www.sintef.no/home/Press-Room/Research-News/Big-Data--for-better-or-worse/>
- “From now until 2020, the digital universe will about double every two years.”  
IDC (2012); <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>

# Big Data

## Redes Sociais

- “Every minute of every day we create: More than 204 million email messages, over 2 million Google search queries, 48 hours of new YouTube videos, 684.000 bits of content shared on Facebook, \$272.000 spent on e-commerce”  
Datamation (2013); <http://www.datamation.com/applications/big-data-analytics-overview.html>
- “one in four people worldwide use social media” (11% Facebook)  
eMarketer (2013); <http://www.emarketer.com/Article/Social-Networking-Reaches-Nearly-One-Four-Around-World/1009976>
- “80 percent of companies use social media for recruitment; 95 percent of those companies use LinkedIn.”  
Psychology Today (2009); <http://www.psychologytoday.com/blog/wired-success/200908/using-social-networking-recruitment-and-training>

## Dispositivos conectados (Internet of Things - IoT)

- “Since 2013 650 million new physical objects have come online; ... 10 percent of automobiles became connected; ... In 2015, all of these things will double again.”  
Gartner (2014) <http://www.gartner.com/newsroom/id/2865519>
- “By the end of 2014, the number of mobile-connected devices will exceed the number of people on earth ” (1.4 in 2018)  
Cisco (2013); [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white\\_paper\\_c11-520862.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html)
- “Worldwide internet traffic increased by a factor of 12 between 2006 and 2011”  
Cisco (2013); [http://www.cisco.com/c/dam/assets/sol/sp/vni/forecast\\_highlights\\_mobile/index.html](http://www.cisco.com/c/dam/assets/sol/sp/vni/forecast_highlights_mobile/index.html)

# Big Data

## Impacto no Mercado

- “IDC forecasts that the Big Data technology and services market will grow at a 27% compound annual growth rate (CAGR) to \$32.4 billion through 2017 - or about 6 times the growth rate of the overall ICT market.”  
IDC (2013); <http://www.idc.com/getdoc.jsp?containerId=prUS24542113>
- “By 2015, 4.4 million IT jobs globally will be created to support big data, generating 1.9 million IT jobs in the United States”  
Gartner (2012); <http://www.gartner.com/newsroom/id/2207915>
- “\$300 billion potential annual value to US healthcare ...  
\$250 billion potential annual value to Europe’s Public Sector ...  
60% potential increase in retailers operating margin ...”  
McKinsey (2012);  
[http://www.mckinsey.com/Insights/MGI/Research/Technology\\_and\\_Innovation/Big\\_data\\_The\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation)





# Big Data

*Estamos no ponto de partida*

The number 100 represents the peak search interest)



Source: Google Trends, April 2013

# Big Data

## Consenso

The logo for the journal Nature, featuring the word "nature" in a white, lowercase, serif font on a dark red rectangular background.

Big data: The next Google (Set 2008)

The logo for The Economist, featuring the words "The Economist" in a white, serif font on a red rectangular background.

The Data Deluge (Fev 2010)

The logo for McKinsey & Company, featuring the text "McKinsey & Company" in a white, serif font on a dark blue rectangular background.

Big data: The next frontier for innovation, competition, and productivity (Mai 2011)

The logo for Forbes magazine, featuring the word "Forbes" in a large, blue, serif font.

Big Data Trends (Jul 2012)

The logo for Fast Company, featuring the words "FAST COMPANY" in a blue, bold, sans-serif font.The logo for WIRED magazine, featuring the word "WIRED" in a white, bold, sans-serif font on a blue rectangular background.The logo for BusinessWeek, featuring the word "BusinessWeek" in a white, serif font on a red rectangular background.

**Harvard  
Business  
Review**

Big Data: The Management Revolution (Out 2012)

The logo for MIT Sloan Management Review, featuring the words "MIT Sloan" in a large, bold, sans-serif font, with "Management Review" in a smaller, grey, sans-serif font below it.

Strength in Numbers: How Does Data-Driven Decision making Affect Firm Performance? Brynjolfsson, Hitt, Kim (Abr 2011)

The logo for The Aspen Institute, featuring the text "The Aspen Institute" in a serif font, with a blue leaf icon to the right.The logo for Stanford Social Innovation Review, featuring the text "STANFORD SOCIAL INNOVATION review" in a serif font, with "review" in a smaller, italicized font.

# Big Data

## *Fatores de desenvolvimento*

### **1. Changing data types**

*The growth in data is fueled by largely unstructured data from websites and machine-generated data from an exploding number of sensors*

### **2. Technology advances**

*Organizations are storing and analyzing more data because they can*

### **3. Insourcing and outsourcing**

*But as the size and importance of corporate e-commerce channels have increased, many are now eager to insource this data to gain greater insights about customers (data as an asset)*

### **4. Developers discover data**

*"We are the beginning of an amazing world of data driven applications. It's up to us to shape the world."—Tim O'Reilly*

# Big Data

*Fatores de desenvolvimento*

**Fontes:** Pessoas, Ciência, Organizações, Máquinas

## Externos

- Avanços nas Comunicações – Internet/Web
- Redes Sociais, Mobilidade
- “Internet of Things”
- Novo modelo de consumidor
- Novos modelos de negócio (digitalização)

## Internos

- Novas tecnologias
- Gestão da Informação e Análise Estatística
- Transformação e Crescimento do negócio

# Big Data

*Definição (não há consenso)*

**Gartner:** High-volume, high-velocity and high-variety of information assets that demand cost effective, innovative forms of information processing for enhanced insight and decision making

**IBM:** Datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze

**NY Times:** Shorthand for advancing trends in technology that open the door to a new approach to understanding the world and making decisions

**McKinsey:** Large pools of data that can be brought together and analyzed to discern patterns and make better decisions

**Economist Intelligence Unit:** the recent wave of electronic information produced in greater volume by a growing number of sources (i.e. not just data collected by a particular organization in the course of normal business)

# Big Data

## *Definição*

**big data** *n.* *Computing* (also with capital initials) data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges; (also) the branch of computing involving such data.

Junho 2013

# Big Data

## *Definição*

Application Delivery Strategies



Date: 6 February 2001

File: 949

Author: Doug Laney

**3D Data Management: Controlling Data Volume, Velocity, and Variety.** Current business conditions and mediums are pushing traditional data management principles to their limits, giving rise to novel, more formalized approaches.

***META Trend: During 2001/02, leading enterprises will increasingly use a centralized data warehouse to define a common business vocabulary that improves internal and external collaboration. Through 2003/04, data quality and integration woes will be tempered by data profiling technologies (for generating metadata, consolidated schemas, and integration logic) and information logistics agents. By 2005/06, data, document, and knowledge management will coalesce, driven by schema-agnostic indexing strategies and portal maturity.***



# Big Data

## *Definição*

### **Volume**

- Machine data
- Application logs
- Clickstream logs
- External data
- 3rd party data
- Emails
- Contracts
- Geo-spacial data (GIS)

### **Velocity**

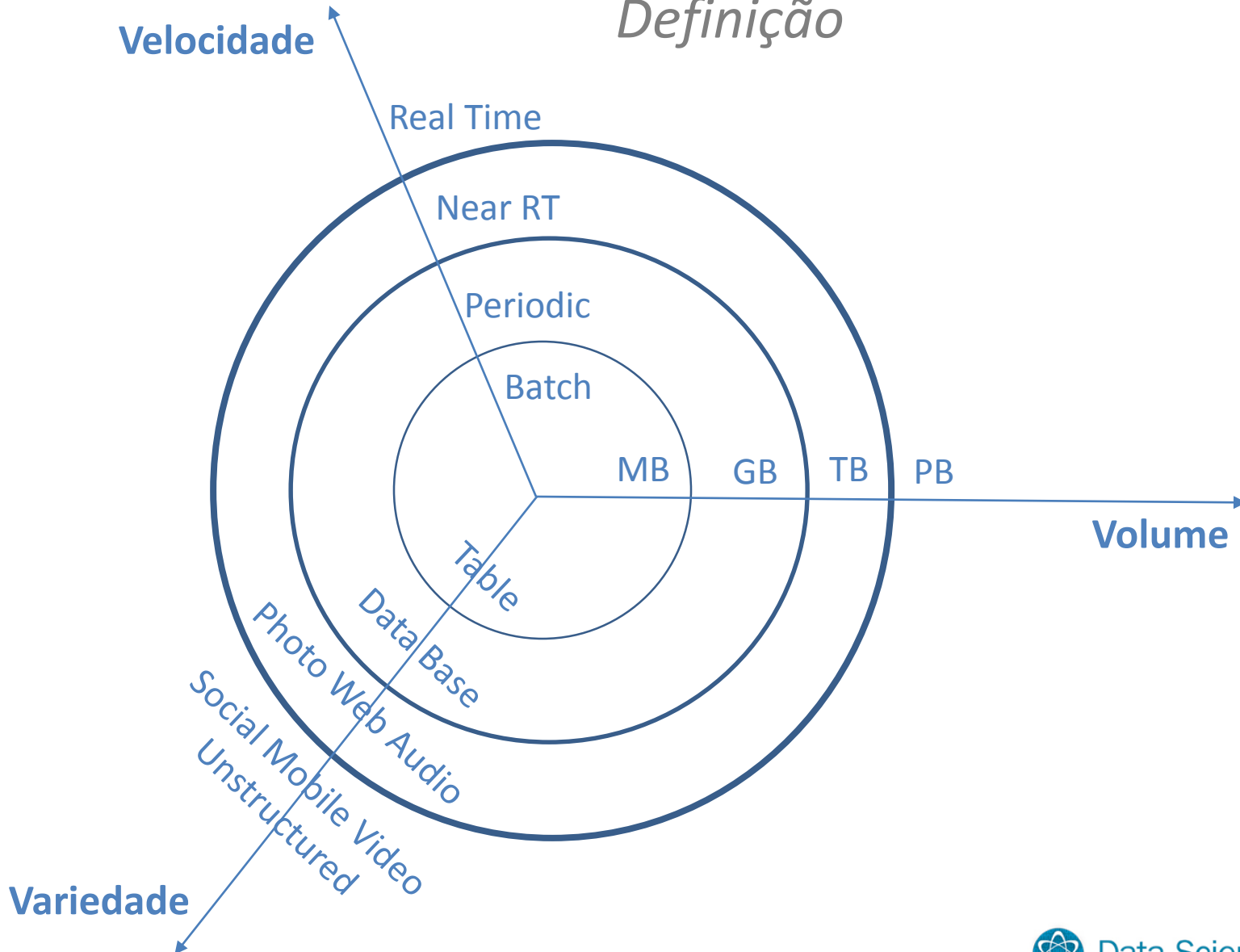
- Social media
- Sensor data
- Mobile networks

### **Variety**

- Data Bases
- Devices
- Applications (API's, SOA, Logs)
- Documents
- Video

# Big Data

## *Definição*



# Big Data

## Exemplos



### Volume

- A Boeing 737 engine generates 10 tb of data per 30 minute in-flight
- The 2 engine 737 will generate 200 tb of data on a 5 hour flight
- Assuming this plane is active for 300 days a year this machine will generate around 60 pb of data a year
- Per Wikipedia Scandinavian Airlines has around 90 Boeing 737 in its current fleet
- *Fun fact: Worldwide, on average about 1,000 737s are in the air at all times*



### Velocity

- The Lotus Formula 1 race car has 240 sensors which registers around 25mb of data per lap
- A race lap is often completed in less than 2 minutes
- The data needs to be collected and analyzed so corrective action can be taken to optimize performance
- Outside of race events, using car sensor data to optimize performance is highly valuable since each lap driven cost around \$450.000
- *Fun fact: Kimi Raikkonen (a world champ in 2007) responding to input from his engineer (car's head mechanic): " Leave me alone, I know how to drive!"*

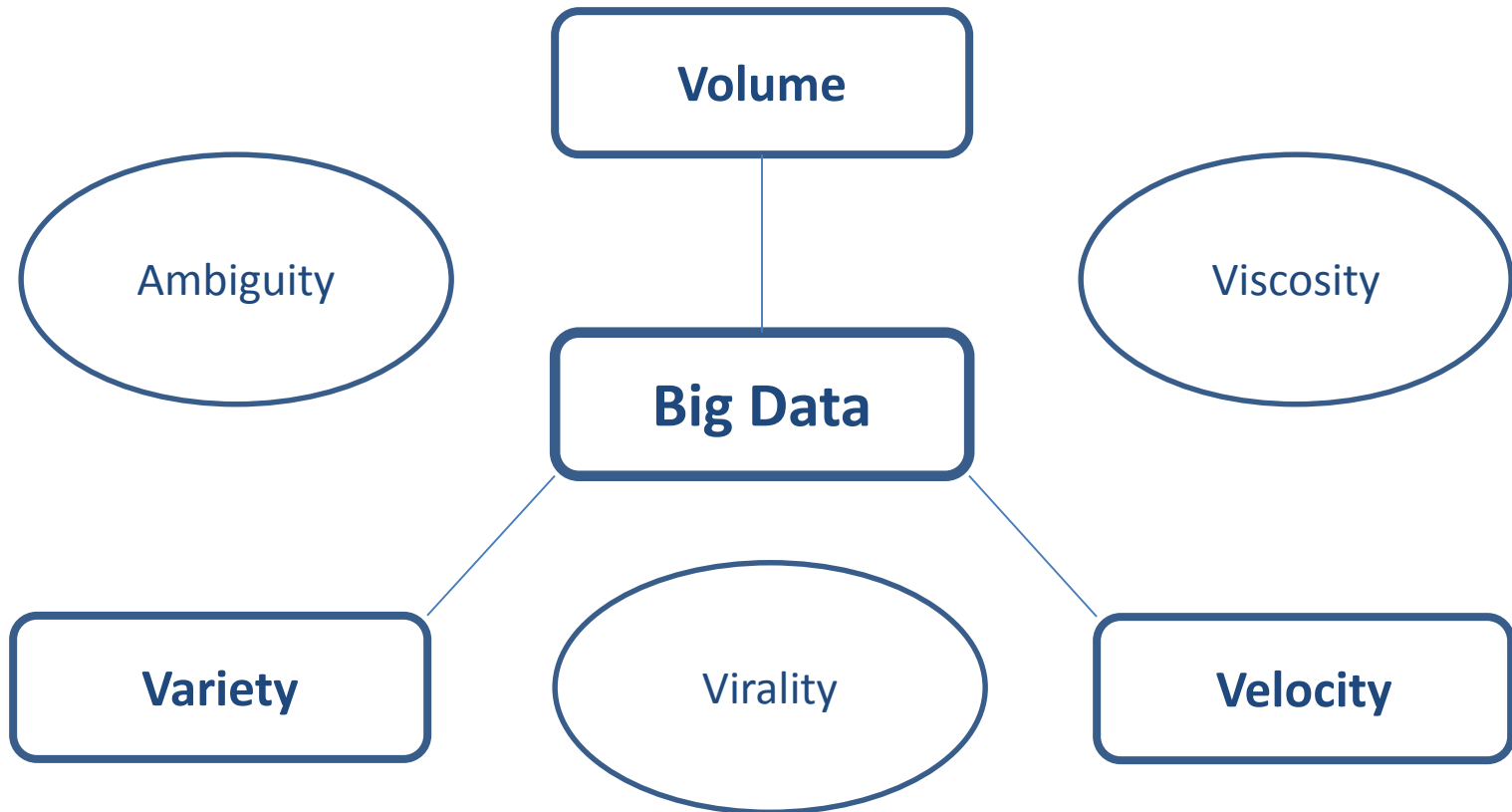


### Variety

- Recent news stories documented that the NSA captures a wide variety of the data generated by humans
- Details of phone records and conversations
- Activity on media and sites such as MSN, Twitter, Google and Facebook are stored and analyzed
- This means that the NSA has to deal with massive amounts of both structured and unstructured data
- *Fun fact: In 2009, German Malte Spitz sued his mobile company over access to his mobile records. Combining this with publically available posts on Twitter and Facebook it has been possible to map out and document most of Malte Spitz' life - <http://www.zeit.de/datenschutz/malte-spitz-data-retention/>*

# Big Data

*Para além dos 3 V's*



# Big Data

*Para além dos 3 V's*

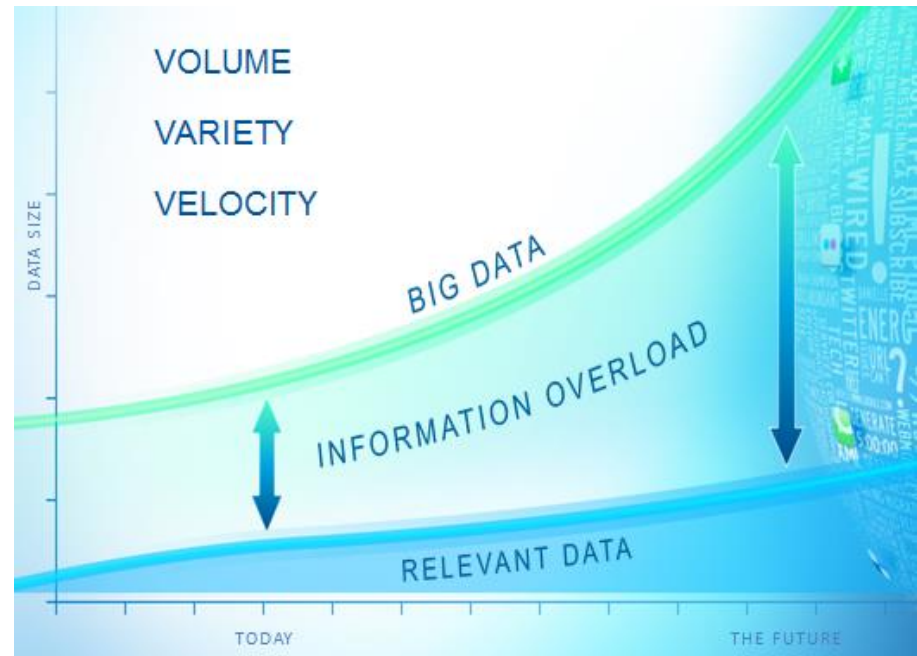
- Validade
- Relevância
- ...

- **Complexidade**

- Ambiente externo

- **Valor**

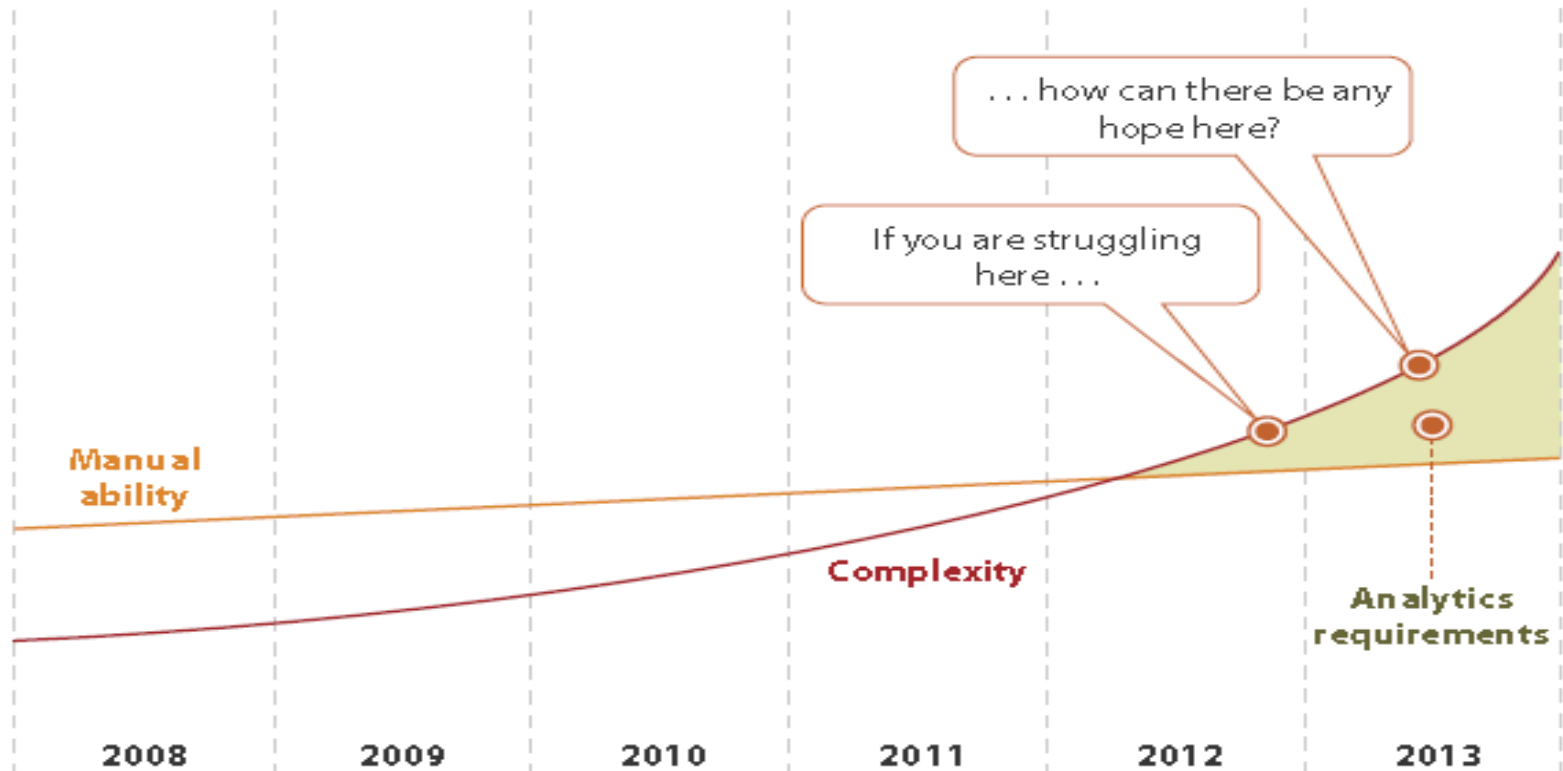
- Ambiente interno
- Tecnologia, arquitetura, processos, competências



# Big Data

## Complexidade

Figure 1 Complexity Has Exceeded Human Capacity



# Big Data

## *Complexidade*

**1- Gerir a informação como um activo estratégico**

Implementar uma abordagem à gestão de informação que resulte numa diferenciação competitiva e criação de valor

**2- Sistematizar processos utilizando técnicas de gestão e análise de informação**

Utilizar técnicas de gestão e análise de informação de forma sistemática e abrangente

Digitalizar e automatizar processos de suporte ao negócio e tomada de decisão

# Big Data

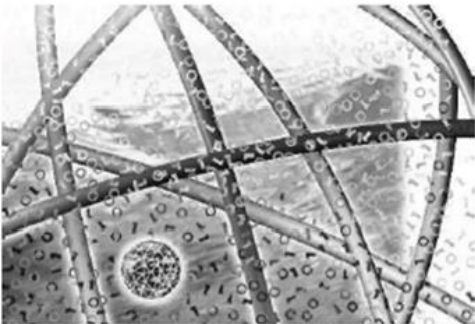
## Valor – Performance

### Strength in Numbers:

#### How Does Data-Driven Decisionmaking Affect Firm Performance?

Erik Brynjolfsson, MIT and NBER  
Lorin Hitt, University of Pennsylvania  
Heekyung Kim, MIT

### Data Driven Decision-makers are Winning



✓ Data-Driven Decision-makers:  
4% higher productivity

✓ Data-Driven Decision-makers:  
6% greater profitability and 50%  
higher market value from IT

Source: Brynjolfsson, Hitt and Kim, 2011



Big data: The next frontier  
for innovation, competition,  
and productivity

McKinsey&Company





# Big Data

*Valor – Transformação*

## Emerging business models (sector view)

Industries are using Big Data to transform business models and to improve performance in many areas.



**Forbes**

## BIG DATA TRANSFORMING BUSINESS

Retail      Fraud Management  
 IT      Billing  
 Online Marketing      Law Enforcement  
 Customer Service      Medical Research  
 Operations Management

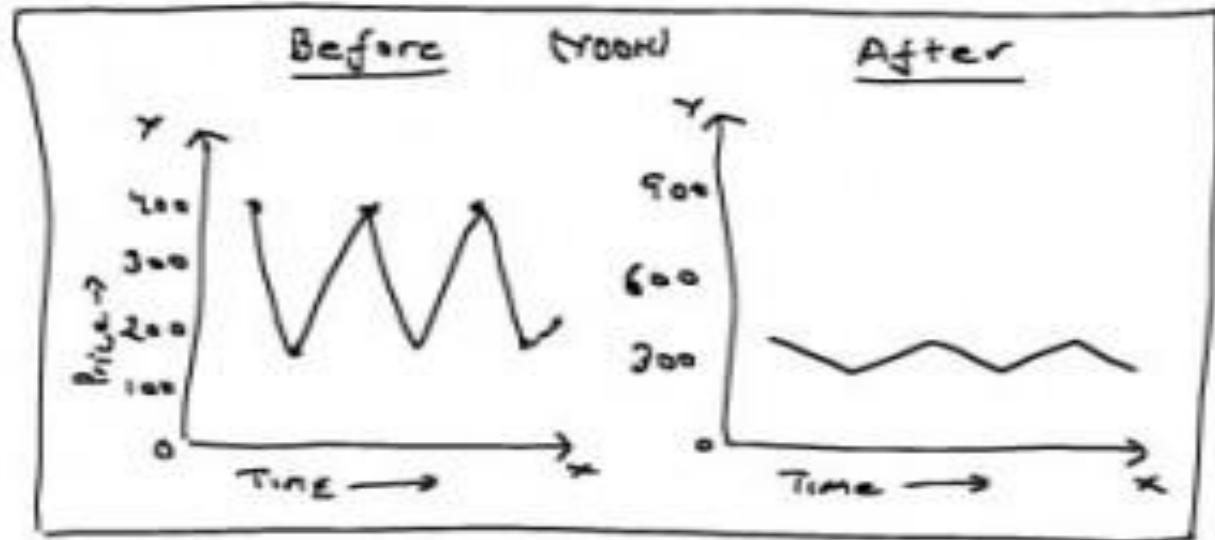
“because of big data, managers can measure, and hence know, radically more about their businesses, and directly translate that knowledge into improved decision making and performance”

Big Data: The Management Revolution  
 HBR, 2012

# **TECNOLOGIAS E ARQUITETURA**

# Tecnologias e Arquitetura

*Mais do mesmo?*



As you can see, we were finally able to control the volatility in our stock price. This was done by carefully increasing the Y-axis intervals.

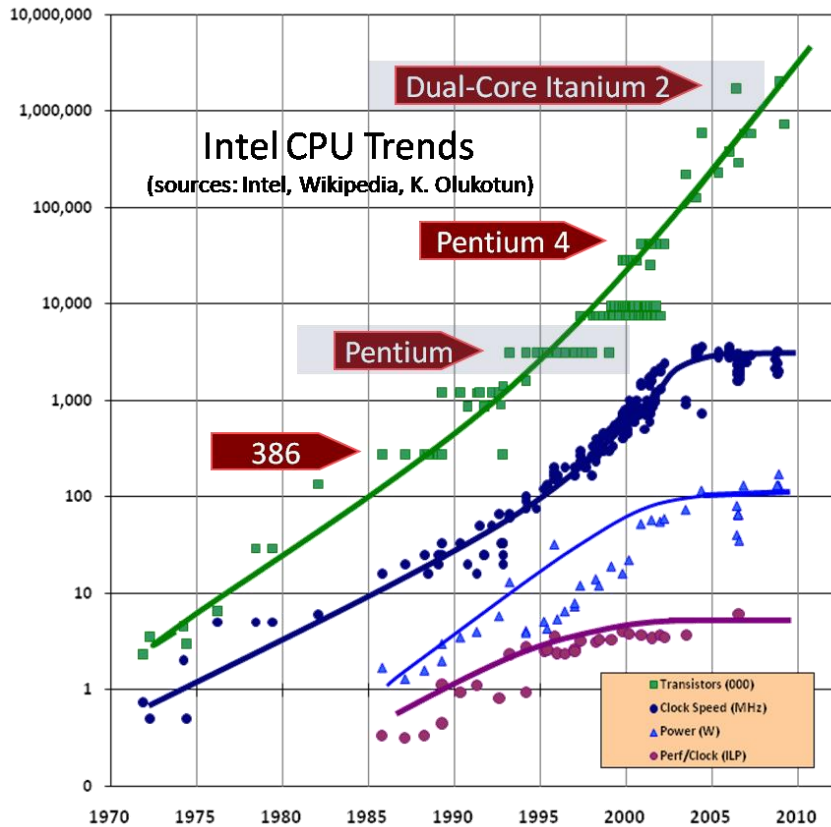
# Tecnologias

*As tecnologias tradicionais não funcionam*

- São grandes volumes de dados
- É variedade e velocidade, claro!
- São dados externos e não estruturados
- Seleção e filtragem (para valor)
- Foco no valor, não na transação
- Experimentação e inovação
- Muitas vezes não sabemos à partida o que procurar (as perguntas)

# Tecnologias

## Technology Push – A tempestade perfeita



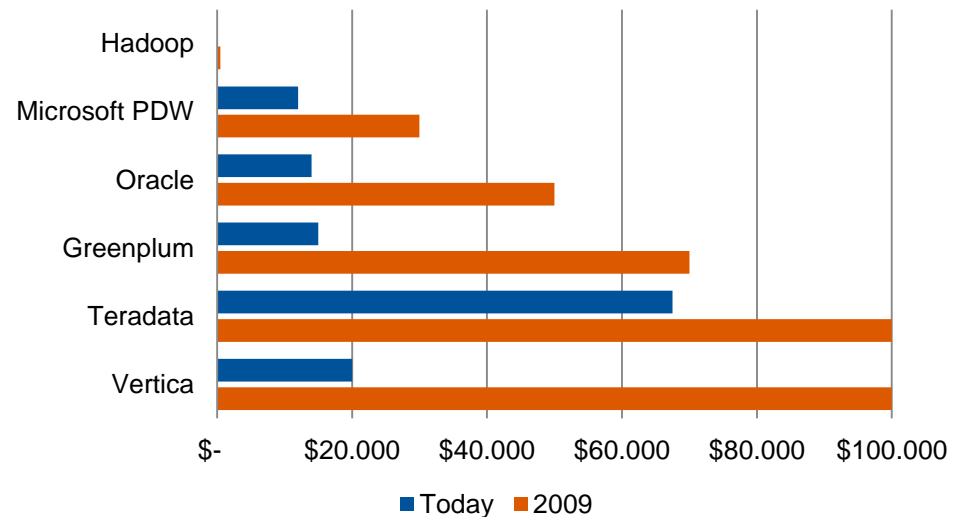
Source: Herb Sutter, Dr. Dobb's Journal

Processamento Paralelo Massivo

### Cost of Storage, Memory, Computing

- In 2000 a GB of **Disk** \$11 today < \$0.05
- In 2000 a GB of **Ram** \$1100 today < \$5

Source: <http://www.statisticbrain.com/>



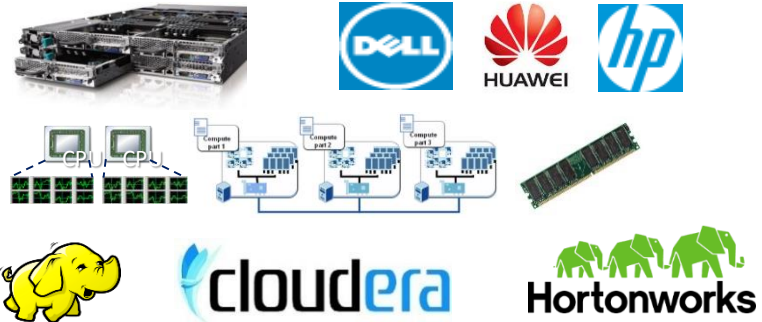
Source: Jack Norris, MapR

Processamento em Memória

# Tecnologias

- **Sistemas de armazenamento e processamento**

- Commodity Hardware
- MPP, Grid, In-memory
- Hadoop (e NoSQL)



- **Sistemas Analíticos**

- In-database Analytics
- In-memory Analytics



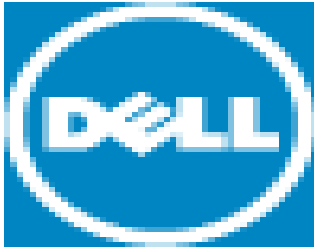
- **Novas soluções BI**

- Analytics
- Visualização, Self-service BI



# Tecnologias

*Commodity Hardware*



# Tecnologias

*MPP, Grid, In-memory*

## Processamento Paralelo Massivo

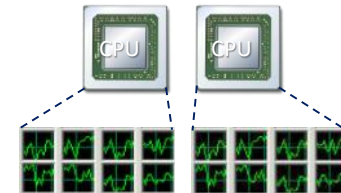
### - MPP

Tira proveito de todos os processadores em paralelo



## Distribuição de Processos - Grid (analíticos e outros)

Explora completamente todos os recursos de múltiplos blades/nodes: RAM e Processadores



## In-Memory

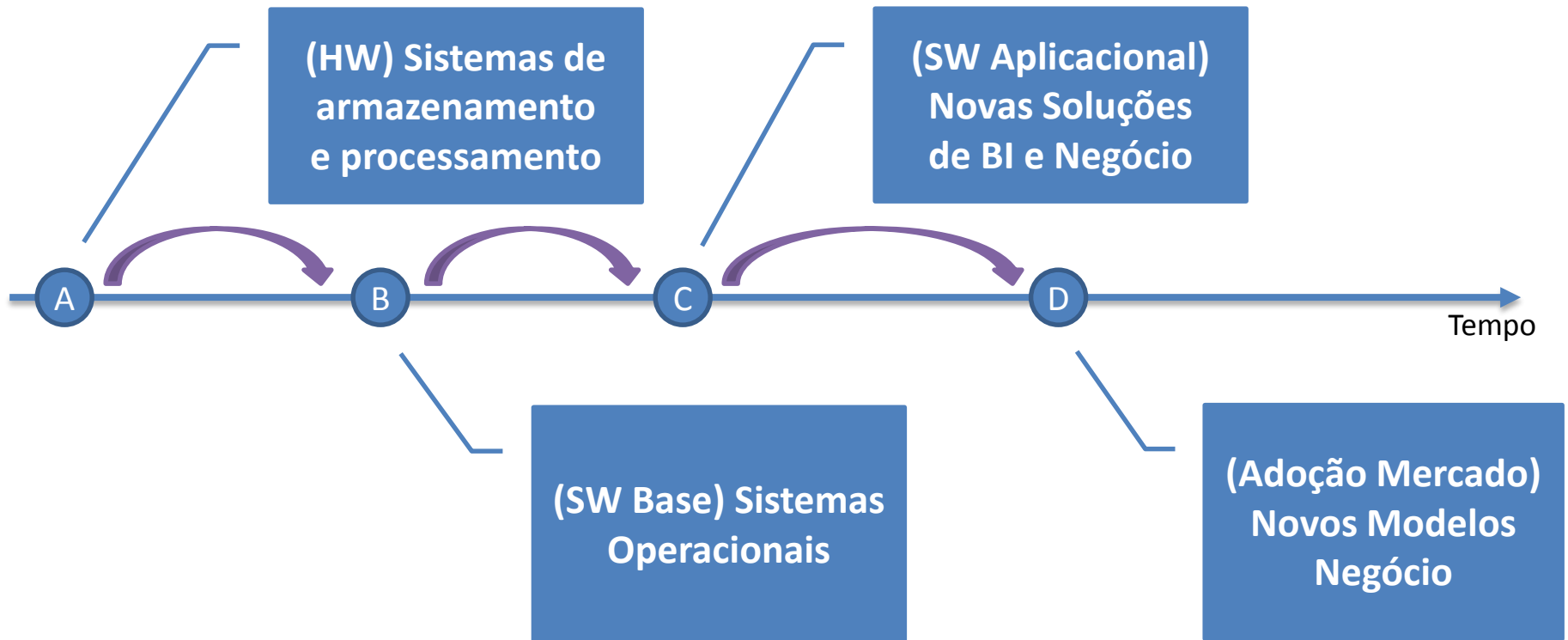
Permite acesso ultra-rápido aos dados





# Tecnologias

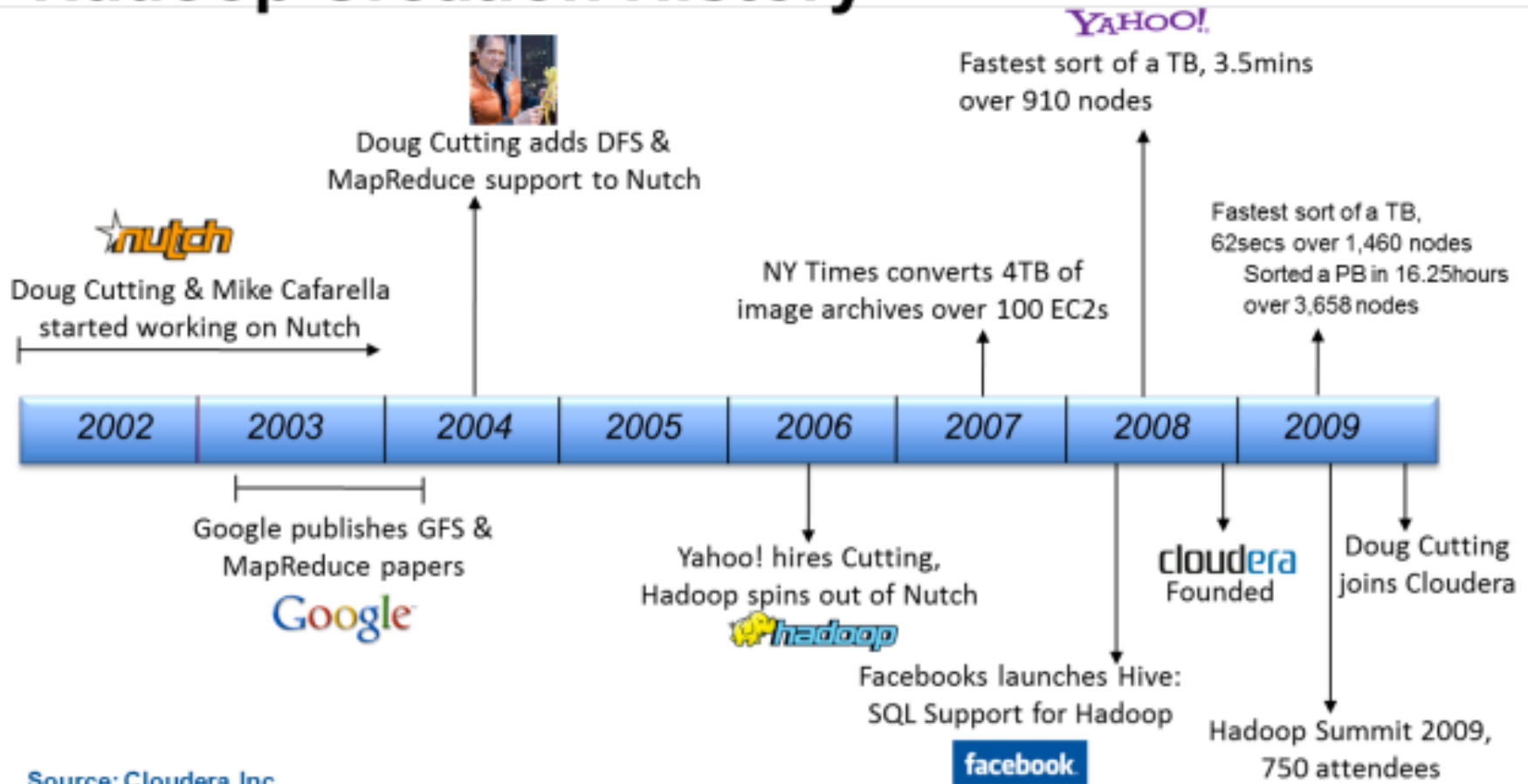
*Quanto tempo durou a Revolução Industrial? (1760-1850)*



# Tecnologias

*Hadoop – Inovação (aberta)*

## Hadoop Creation History



Source: Cloudera, Inc.

# Tecnologias

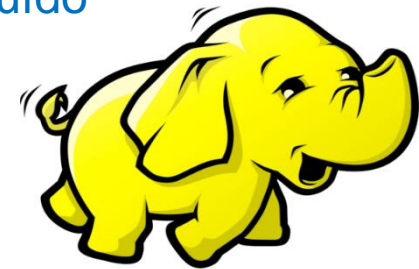
## *Hadoop – Tecnologia disruptiva?*

“Hadoop is one way of using a set of **cheap computers** to **store** an **enormous amount of data** and then to **process** that data in **parallel**.”

Keith Wiley: <http://escience.washington.edu/get-help-now/what-hadoop>

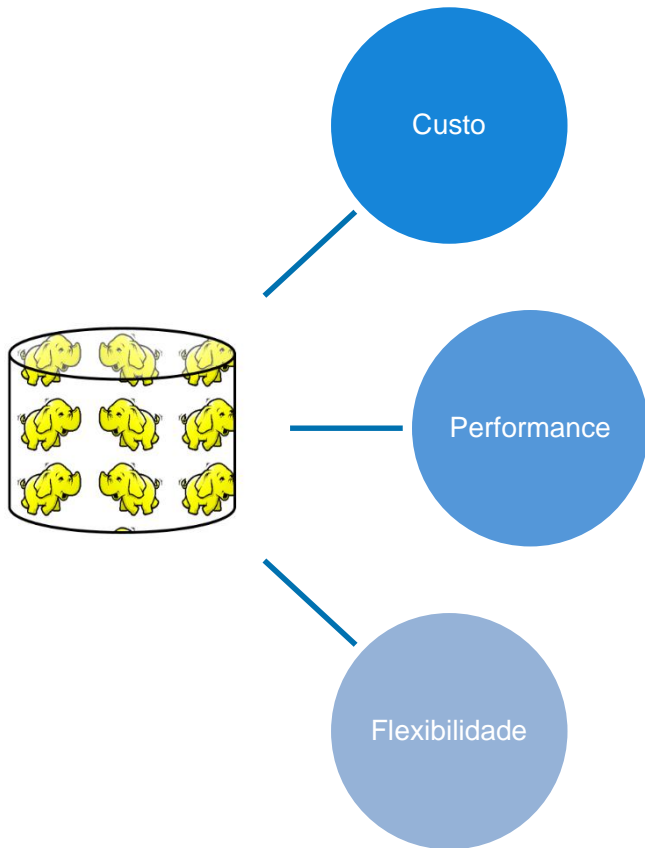
### Plataforma de armazenamento E processamento paralelo distribuído

- Grandes volumes de dados em ambiente distribuído
- Utilização de “*commodity hardware*”
- Framework de código aberto



# Tecnologias

## *Hadoop - Vantagens*



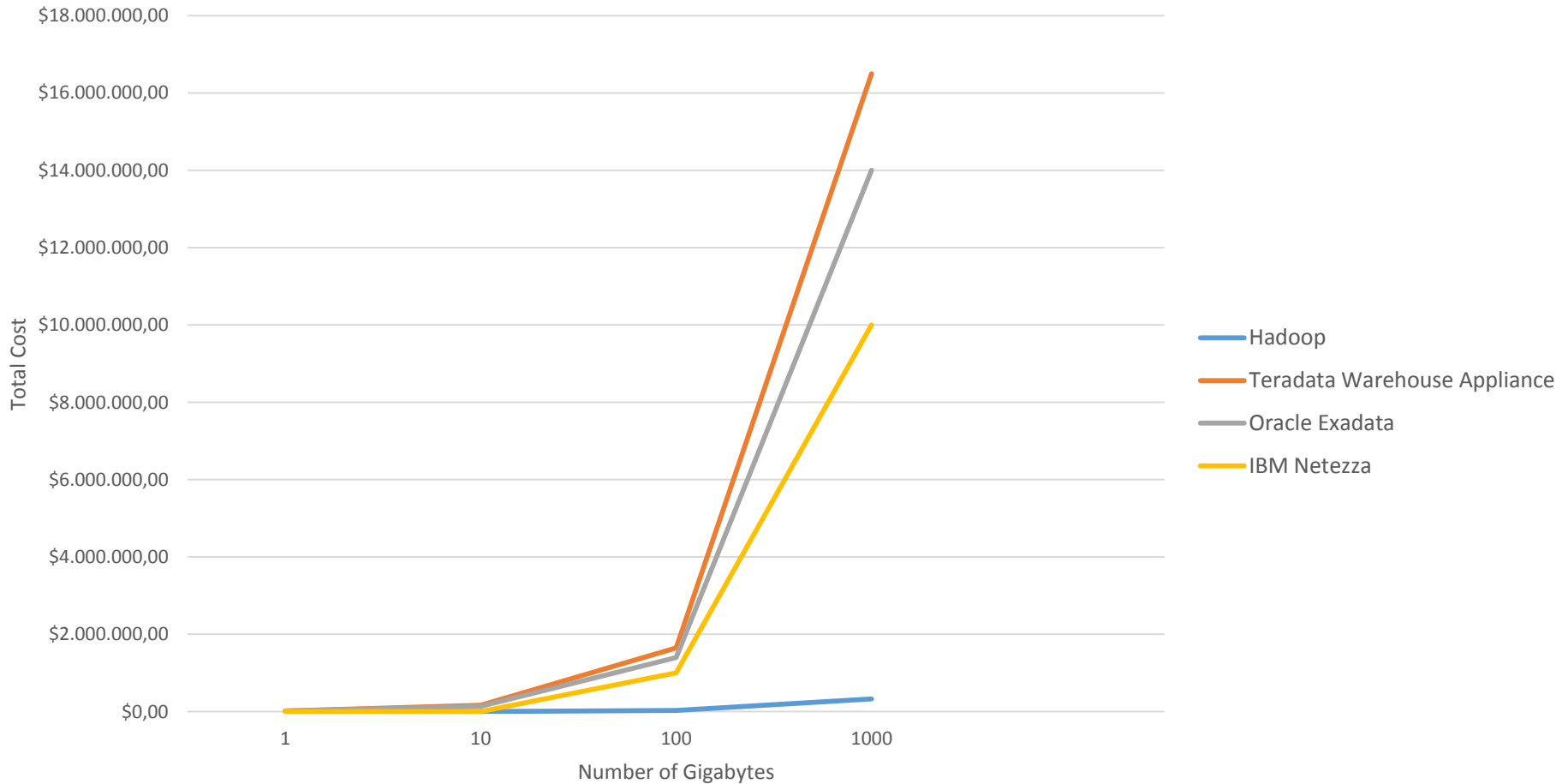
- Base de dados de **baixo custo** baseado commodity HW
- Escalabilidade: começar pequeno e **crescimento linear**
- Framework de **código aberto**

- Processamento massivo em **paralelo**
- Funcionalidades nativas para **tolerância a falhas**
- Rapidez no **processamento analítico** e **carregamento** dos dados independentemente da sua estrutura

- Facilidade de acesso a **novas fontes de dados**
- **Tempo de modelização** reduzido ou nulo
- **Complementaridade à EDW** para carregamento dos dados históricos, ETL e processamento analítico

# Tecnologias

## *Hadoop - Vantagens*



# Tecnologias

## *Hadoop - Vantagens*

Cloudera's Hadoop distribution is **like a data warehouse**, but it can store **more data, more kinds** of data, and perform **more flexible** analyses.

Also, it's **open source** so it's orders of magnitude **more economical** than data warehouse solutions.

Traditional Data Warehouse

Add 100 TB =

**\$2M TO \$10M**

in incremental spend

With Cloudera

Add 100 TB =

**\$200K**

1/10th the cost of legacy systems

# Tecnologias

## *Hadoop – Luta pela posição dominante no mercado*



# Tecnologias

*Hadoop – Luta pela posição dominante no mercado*

**GLOBAL HADOOP MARKET**  
**\$50.24 BILLION IN 2020**



**GROWING AT CAGR 58.2% (2013-2020)**



**NORTH AMERICA**

**Asia Pacific  
Europe  
RoW**

**HIGHEST REVENUE  
GENERATING GEOGRAPHY  
\$25.85 BILLION (2020)**

**GLOBAL HADOOP APPLICATION  
SOFTWARE MARKET**



**APPLICATION  
SOFTWARE MARKET**



**PACKAGED  
SOFTWARE MARKET**

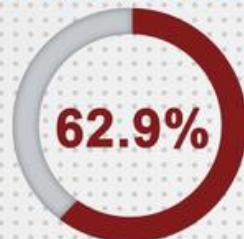


**PERFORMANCE  
MONITORING  
SOFTWARE MARKET**



**MANAGEMENT  
SOFTWARE MARKET**

**FASTEST GROWING SEGMENT  
EXPECTED TO GROW  
AT A CAGR OF**



**DURING 2012-2020**

**GLOBAL HADOOP  
HARDWARE MARKET**



**SERVERS  
MARKET**

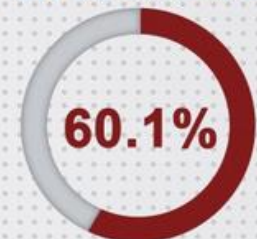


**STORAGE  
MARKET**



**NETWORK AND  
EQUIPMENT MARKET**

**FASTEST GROWING SEGMENT  
EXPECTED TO GROW  
AT A CAGR OF**



**DURING 2012-2020**



# Tecnologias

*Hadoop – Luta pela posição dominante no mercado*

Figure 1. Magic Quadrant for Data Warehouse Database Management Systems



# Gartner®

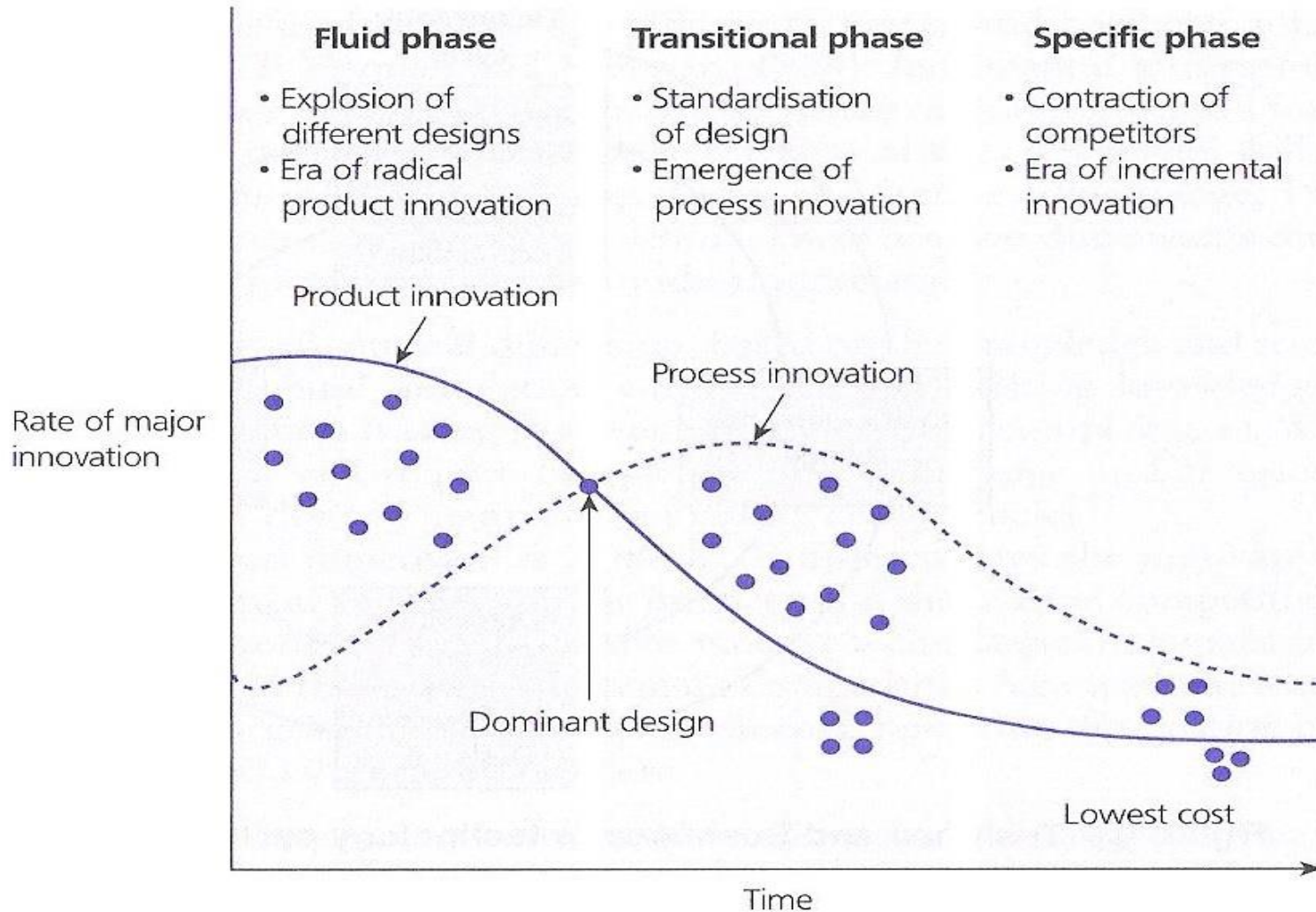
Hadoop no Quadrante Mágico da Gartner para Sistema BD para Data Warehousing (Março 2014)



cloudera

# Tecnologias

## *Hadoop – Luta pela posição dominante no mercado*



**Abernathy and Utterback's three phases of innovation**

Source: Utterback (1994).

# Tecnologías

## *Sistemas Analíticos*

### Statistics

- Binary target & continuous no. predictions
- Linear, Non-Linear, & Mixed Linear modeling

### Data Mining

- Complex relationships
- Tree-based Classification
- Variable Selections

### Text Mining

- Parsing large-scale text collections
- Extract entities
- Auto. stemming & synonym detection

### Forecasting

- Large-scale, multiple hierarchy problems

### Econometrics

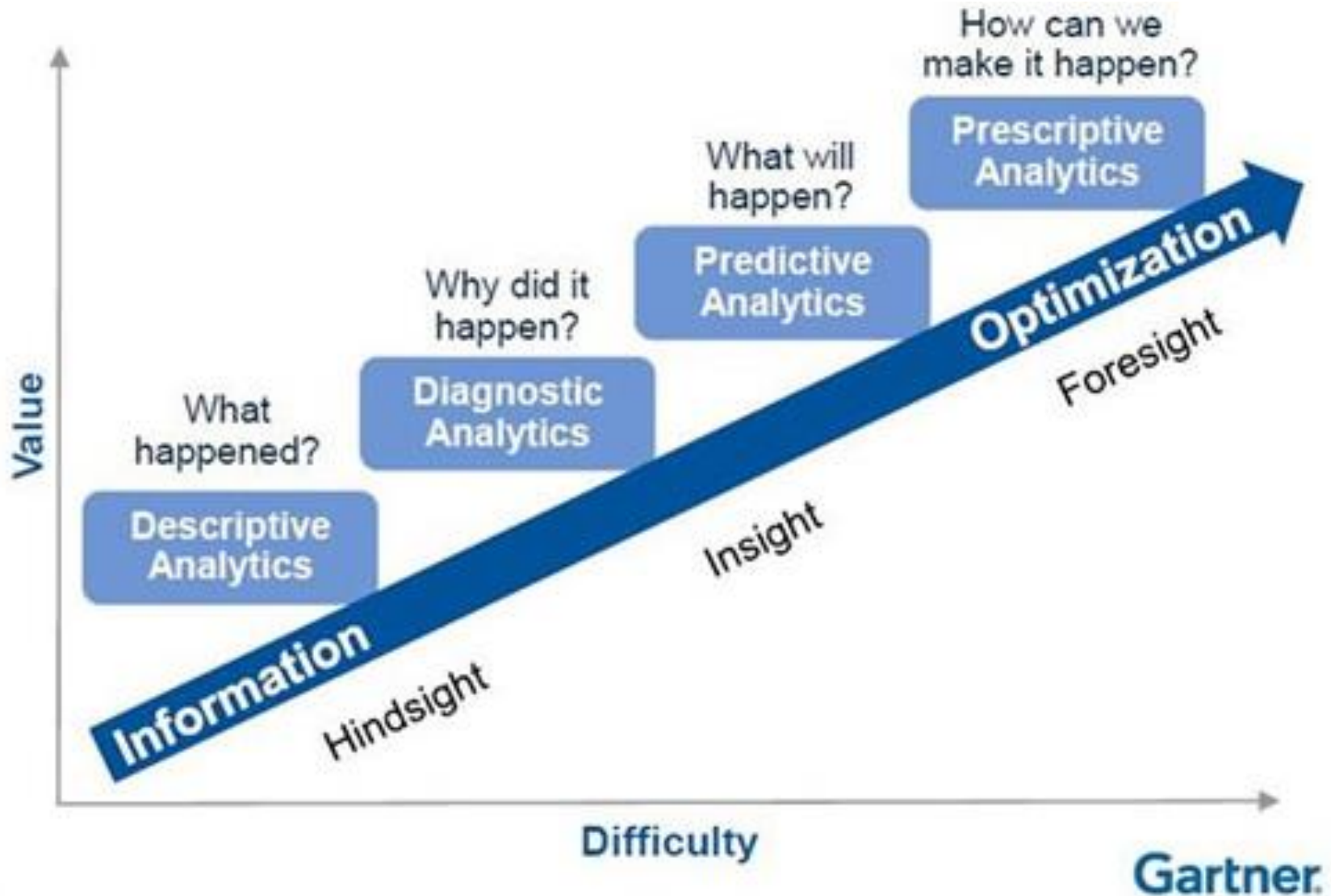
- Probability of events
- Severity of random events

### Optimization

- Local search optimization
- Large-scale linear & mixed integer problems

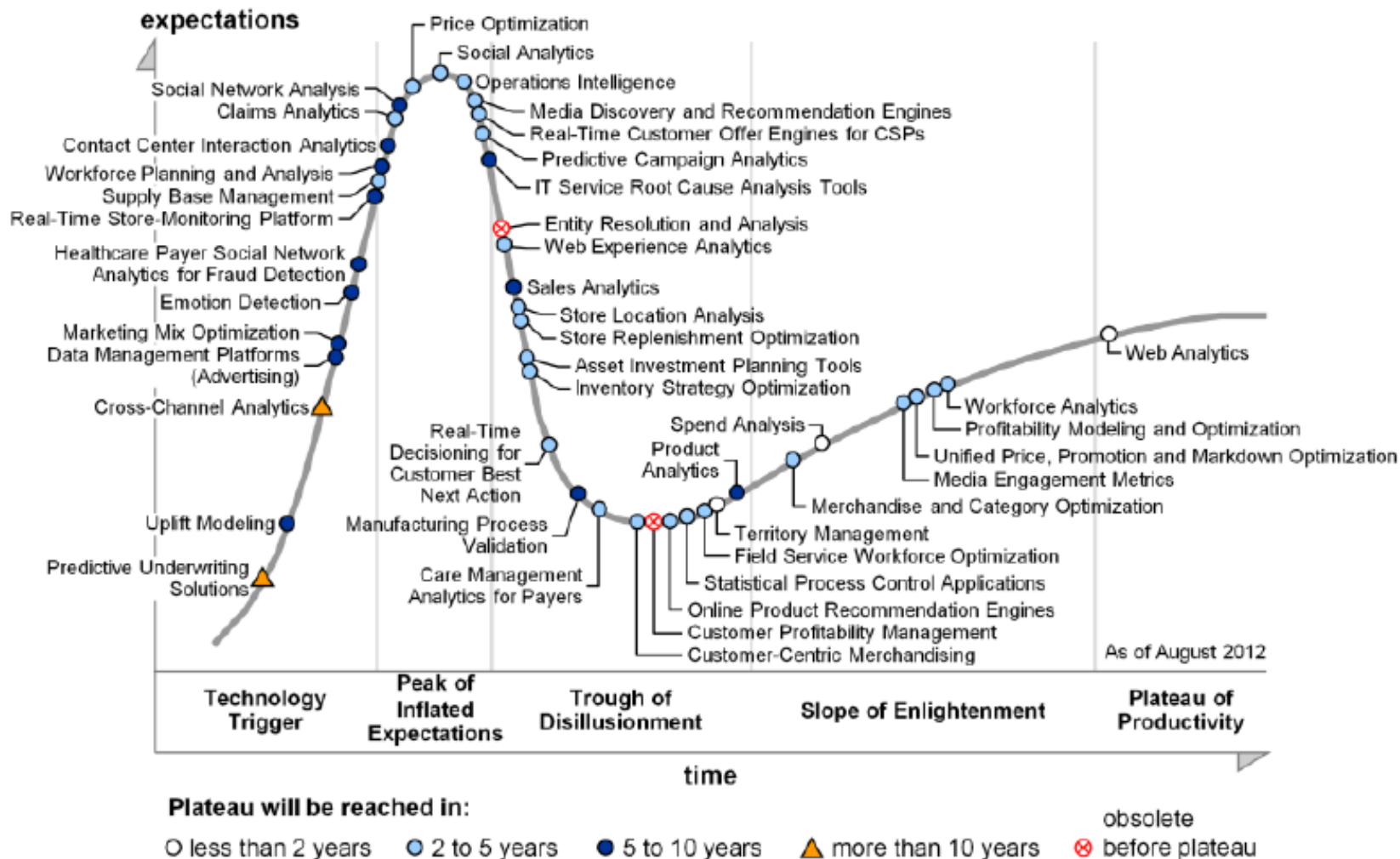
# Tecnologías

## *Sistemas Analíticos - Valor*



# Tecnologias

## Sistemas Analíticos - Maduridade



# Tecnologias

*Novas ferramentas de BI*

Self-service BI

Report & Discover



**REPORTING & MOBILE**

SHARE INSIGHTS WITH EVERYONE EVERYWHERE  
DESKTOP  
WEB & MOBILE



**DATA VISUALIZATION**

**VISUAL ANALYTICS**

SAS ANALYTICS  
POWERFUL VISUALIZATIONS  
AT THE SPEED OF THOUGHT  
RICH VISUAL EXPLORATION  
RAPID DISCOVERY  
DEEPER INSIGHTS  
INTELLIGENT ADAPTIVE VISUALIZATIONS  
ACCELERATE YOUR ANALYTIC LIFECYCLE



**IN-MEMORY**

HIGH PERFORMANCE  
ALL OF YOUR DATA  
ANY VARIETY OF MEASURES  
ULTRA HIGH SPEED  
COST-EFFICIENT SCALABILITY  
FLEXIBLE DATA PREP  
VAST NEW DATA SOURCES  
OVERCOME LIMITS

**APPROACHABLE ANALYTICS**

BROADENS AUDIENCE FOR ANALYTICS  
EASILY EXPLORE AND VISUALIZE ANALYTIC RESULTS  
FAR LESS BURDEN ON IT  
PERVERSIVE, AGILE AND FLEXIBLE

Scalability & Performance

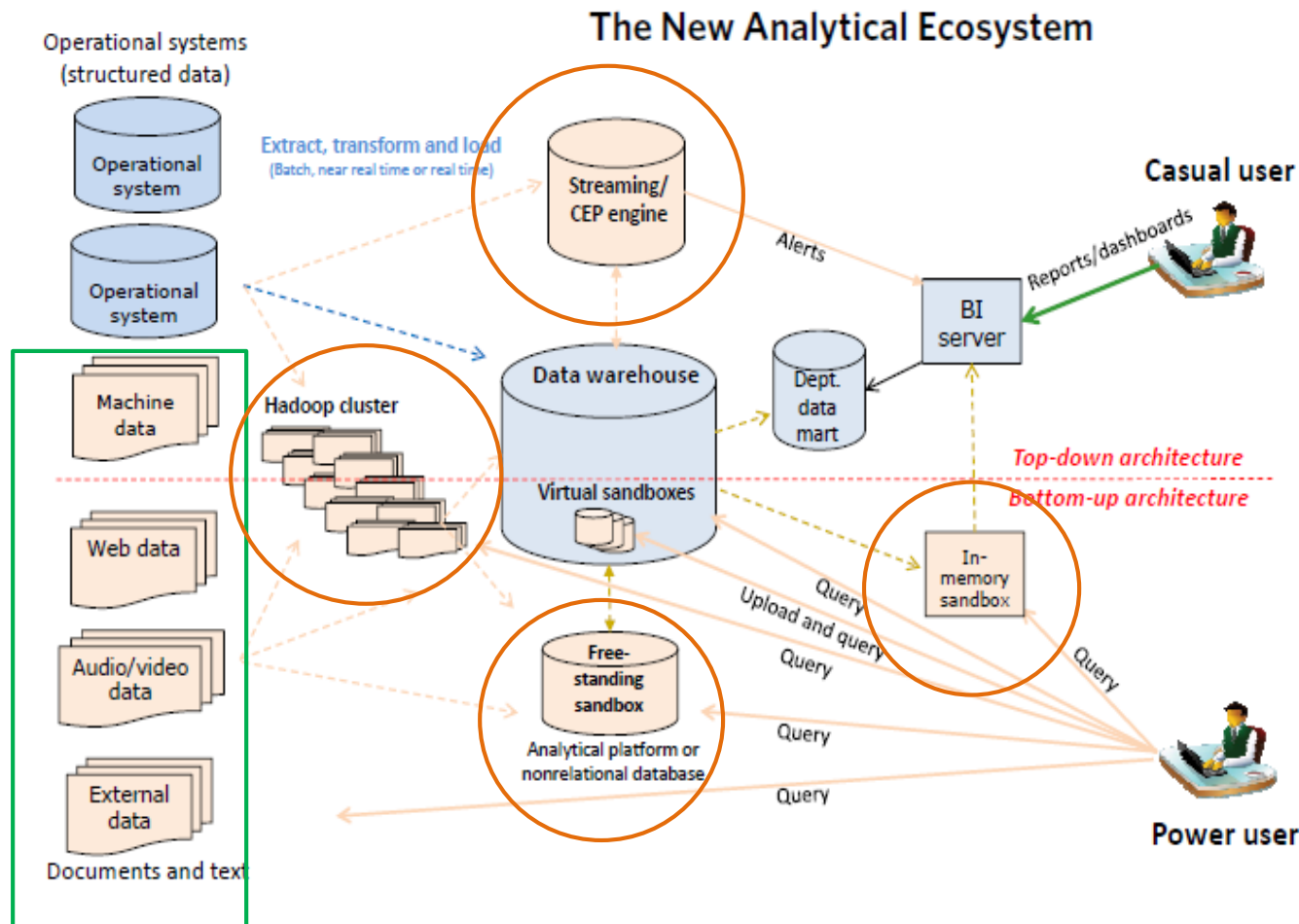
Analytics

# Arquitetura

## Complexa, Híbrida

*The next-generation BI architecture is more analytical, giving power users greater options to access and mix corporate data with their own data via various types of analytical sandboxes. It also brings unstructured and semi-structured data fully into the mix using Hadoop and nonrelational databases.*

Exploiting Big Data  
Wayne Eckerson, 2012



# ADOÇÃO E VANTAGENS



# Adoção e vantagens

## Big data is taking off

Users that have completed at least one project are very satisfied with their initial forays into big data. The vast majority report that they are satisfied with business outcomes and that their big data initiative is meeting their needs.



Of users are fully satisfied with their business outcomes.



Of users report that their implementation is meeting their needs.



Of users believe big data will revolutionize operations the same way the Internet did.



Of users believe big data is very important to their transformation into digital.

 **accenture**

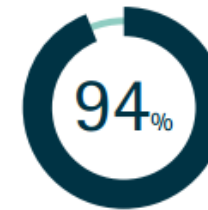
High performance. Delivered.

## Big Success with Big Data

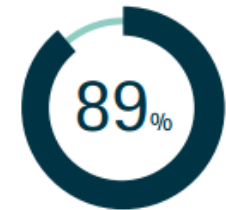
Respondents from organizations that are using big data today report overwhelming satisfaction with their results and see big data as a catalyst for their company's transformation to become a digital enterprise.



## What are users doing with big data?



identifying new sources of revenue



developing new products or services

Accenture (2014)

# Adoção e vantagens

## “Where we at?”

Big Data is no longer a promise nor a trend. Big Data is here and is sparking profound changes in various industries. From a technological point of view, there are already many projects and products that have gained widespread adoption in certain industries. The analysis of all available information is becoming a disruptive element. Just like the Internet, it is a disintermediation factor that is affecting many value chains. The analysis of large volumes of information, from different sources, at high speed, and with unprecedented flexibility can be a differentiating factor for anyone who decides to adopt it.

Source: BBVA Innovation center – Big Data (2013)

# Adoção e vantagens

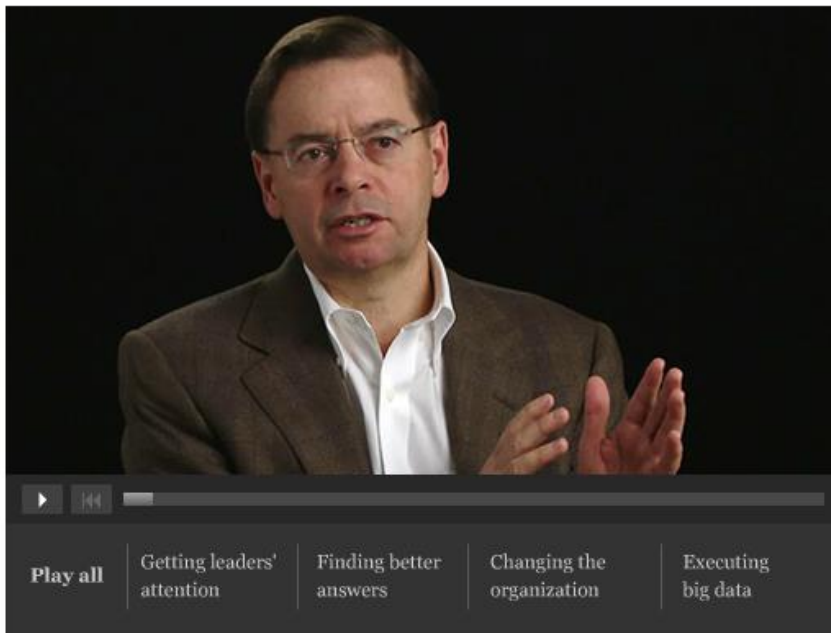
“**The payoff** from joining the big-data and advanced-analytics management revolution is **no longer in doubt** ... they can deliver productivity and profit gains that are **5 to 6 percent higher** than those of the competition.”

McKinsey&Company

## Putting big data and advanced analytics to work

In a video feature, McKinsey director David Court explains how companies can improve their decisions and performance by getting powerful new tools in the hands of frontline managers.

September 2012



### About this content

The material on this page draws on the research and experience of McKinsey consultants and other sources. To learn more about our expertise, please visit the [Business Technology Practice](#), [Consumer Packaged Goods Practice](#), [Marketing & Sales Practice](#), [Retail Practice](#), [High Tech Practice](#).

## key challenges

- Manage your data
- Build analytics models
- New intuitive tools
- Transform your business

# Adoção e vantagens

- Scientific revolution begin with measurement
- Big Data is not just a tech revolution - **it's a Management revolution**
- Data-driven decision makers are better decision-makers
- Data is not knowledge and correlation is not causality
- Measure, experiment, analyze and replicate

# Adoção e vantagens

## Vantagens para a empresa

- Suporte à Decisão (Data Driven Decisions)
- Conhecimento do Cliente e Mercado
- Inovação, DNP
- Novos Modelos Negócio
- Vantagem Competitiva

## Vantagens para a sociedade

- Ciência, Governo, Saúde, Segurança Pública
  - McKinsey - Big data: The next frontier for innovation, competition, and productivity (Maio 2011)
  - UN Global Pulse - Big Data for Development: Challenges & Opportunities (Maio 2012)

# Adoção e vantagens

## *Casos de Uso*

### Potential Use Cases for Big Data Analytics



# Adoção e vantagens

## *Casos de Uso*

- Visão 360º do cliente, experiência do cliente, necessidades do cliente
- Análise e descoberta de novos padrões e comportamentos
- Análise e melhoria das operações
- Otimização do SSD/Data Warehouse
- Suporte a novos modelos de negócio (digitais)
- “Internet of Things”  
(OilGas, Energy, Telco, HealthCare, Home etc)
- Segurança e deteção de fraude

# PESSOAS E PROCESSOS



# Processos e Pessoas

*A necessidade de “Exploração” (aprender com os dados)*

## **Projetos tradicionais de BI**

- Os requisitos de negócio vêm em primeiro lugar - são estáveis
- Centrados nos processos
- São planejadas as conversas com o negócio, definidos os objetivos e vantagens, desenhado o sistema, a arquitetura e finalmente feita a implementação

## **Projetos de Big Data**

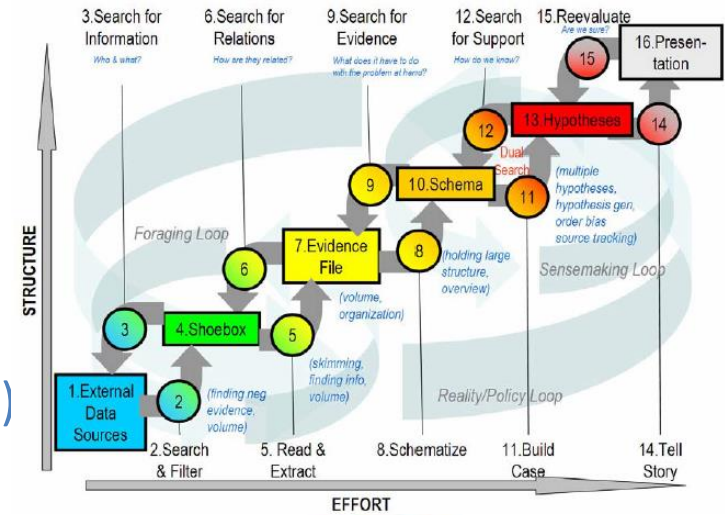
- Não se sabe o suficiente sobre o que se está a passar. Não é possível definir com rigor todos os requisitos de negócio
- Centrados nos dados
- É necessária uma fase de exploração, verificação e aprendizagem
- Mesmo considerando uma ideia inicial, ela tem de ser validada e muitas vezes os resultados obrigam fazes exploração adicionais

# Processos e Pessoas

*Para além dos processos e competências tradicionais*

## Processos para Big Data

- The Sesemaking Process  
Peter Pirolli and Stuart Card (2005)
- Analitical Sand-box (Experimentação)



## Pessoas

- “Data Scientist”
- “Chief Data Officer”
- “Chief Marketing Technologist”?



# Pessoas e Processos

*O Data Scientist*

**“Statistics is the next sexy job”**

Hal Varian – Google Chief Economist

**Data Science:** transformar dados em conhecimento (produtos/iniciativas/modelos) de negócio, utilizando um processo sistemático, rigoroso e fundamentado

**International Association for Statistical Computing**

*Data science is the linkage of traditional statistical methodology, modern computer technology, and the knowledge of domain experts in order to convert data into information and knowledge*

# Pessoas e Processos

## O “Data Scientist”

**Data Scientist:** resolve problemas complexos de dados recorrendo a métodos científicos especializados

### 3 pilares de conhecimento:

- **Conhecimento de Negócio**  
Analista de dados e negócio
- **Conhecimento Técnico**  
Engenharia Informática e Estatística
- **Gestão e Inovação**  
Capacidade de análise, exploração e criatividade negócio  
Capacidade de apresentação e comunicação

**INOVAÇÃO SEGURANÇA E ÉTICA**

# Inovação, segurança e ética

## *Inovação*

- Novas fontes de informação, externas, diversas - medir
- Nova tecnologia, combinada e orientada à gestão da informação e ao utilizador de negócio (vs IT)
- Novas competências, combinadas, dinâmicas
- Experimentação e (novo) conhecimento
- Alteração aos processos e do modelo de negócio
- Replicação, desenvolvimento constante e “realimentação”

*Big Data é um catalisador do processo de inovação (digital)*

# Inovação, segurança e ética

*Segurança e ética*

**A tecnologia permite mas o elemento humano impera!**

**Objetivos de negócio**

**Direitos dos cidadãos**

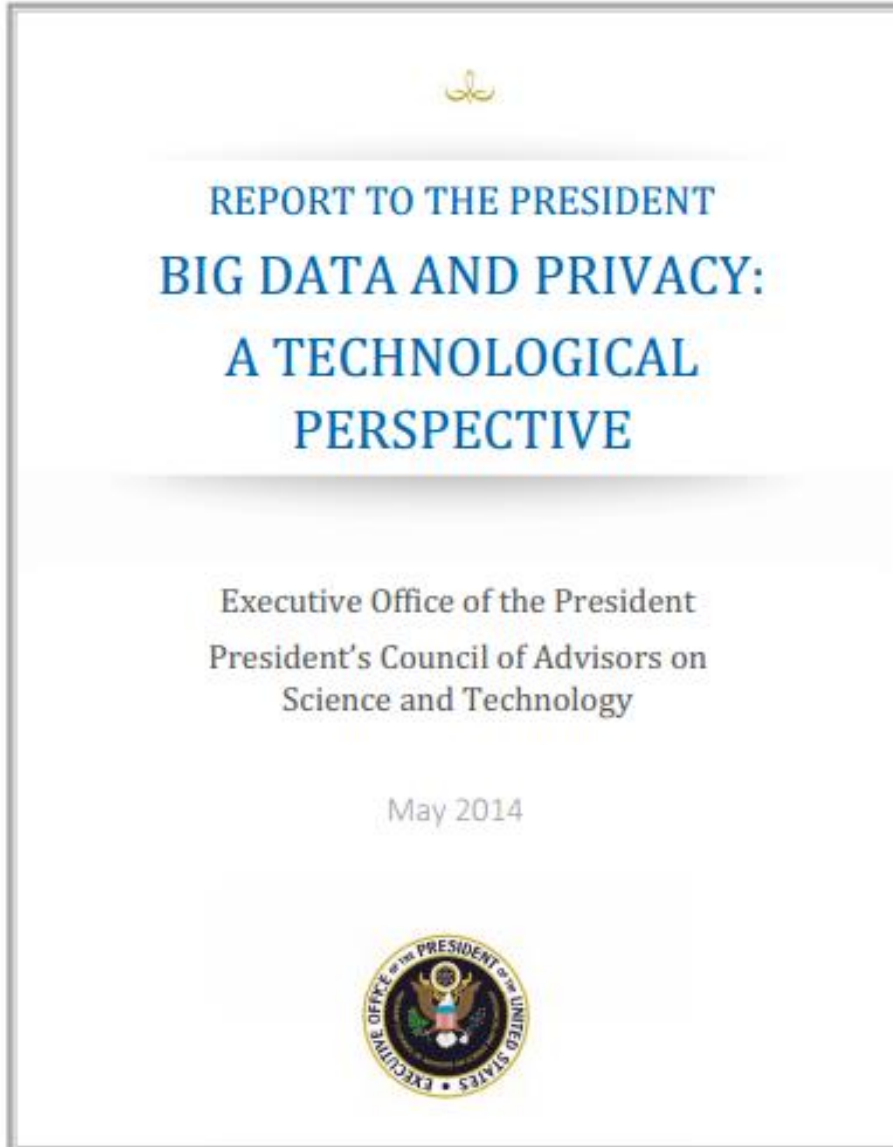
**Legislação**

**Atividades ilícitas**

- Recolha e comercialização de dados
- Segurança e privacidade dos dados, pessoas e bens
- Preferências e personalização
- Direitos e responsabilidades
- Crimes, fraude, abuso

# Inovação, segurança e ética

## *Segurança e ética*



“limiting collection, storage and retention of information is unfeasible”

“this approach is unnecessarily restrictive and could stifle the economic and societal benefits of data analysis”

“Only in some fantasy world do users actually read these notices and understand their implications”

“burden of privacy protection currently falls mainly on consumers”

“Look at uses of data as potential targets of regulations”

“new regulations may be needed”



# CONCLUSÕES

# Conclusões

## *Big Data*

- Inovação que endereça necessidades importantes das **organizações** (otimizar, crescer) baseadas em informação (externa) abundante mas difícil de tratar com a tecnologia incumbente (standard)
- A abordagem tradicional não funciona. Exige novas **tecnologias** (push) e conhecimento gerados muitas vezes (em colaboração) pelo mercado (fora do seu núcleo tradicional de desenvolvimento)
- O principal desafio é a gestão da **complexidade externa** (inf) que implica alterações às competência e funções da empresa
- O foco na **gestão e análise da informação** é crítico (digital) passando esta a ser um **novo ativo** da empresa
- Leva a uma **transformação** dos processos e do modelo de negócio que é **difícil** (e leva tempo a concretizar): processo de decisão e digitalização do negócio
- É um catalisador do processo de **inovação** e representa uma importante vantagem competitiva para a empresas mas traz **novos desafios e preocupações sociais**

# Big Data



**Instituto Superior de Economia e Gestão**

---

UNIVERSIDADE TÉCNICA DE LISBOA