



# XXI GLAIN COSTA RICA 2017

18-19 | Congreso Latinoamericano de Auditoría Interna y Evaluación de Riesgos | Hotel Intercontinental  
Mayo 2017

## **Big Data** como herramienta para la gestión de riesgos en los bancos

**Leandro Fernández Lacalle**

Gerente General - AIS Chile y Región Andina

Organizan



# Introducción o antecedentes

El Big Data está revolucionando aspectos clave de la banca, como son la **Gestión del Riesgo y el Marketing**.

Con el crecimiento explosivo de los datos y la capacidad técnica de **convertir estos datos en conocimiento**, los bancos pueden mejorar:

- ✓ La segmentación de sus clientes
- ✓ Modelos predictivos
- ✓ Efectividad en captación
- ✓ Fidelización de clientes
- ✓ Detección de fraude
- ✓ Estimación de ingresos
- ✓ Etc..



# Agenda

- **Conceptos y mitos del Big Data**
- Impacto del Big Data en distintas industrias
- Desafíos del Big Data: Roles, infraestructura y regulación
- Casos prácticos de aplicación





# ¿Qué es el BigData?

- “Gestión y análisis de **grandes volúmenes** de datos que no pueden ser tratados de forma convencional”
- “Métodos y técnicas para **integrar, almacenar y analizar datos poliestructurados**”
- “High-volume, high-velocity and/or high-variety **information assets** that demand **cost-effective**, innovative forms of information processing that enable enhanced insight, **decision making**, and process automation.”

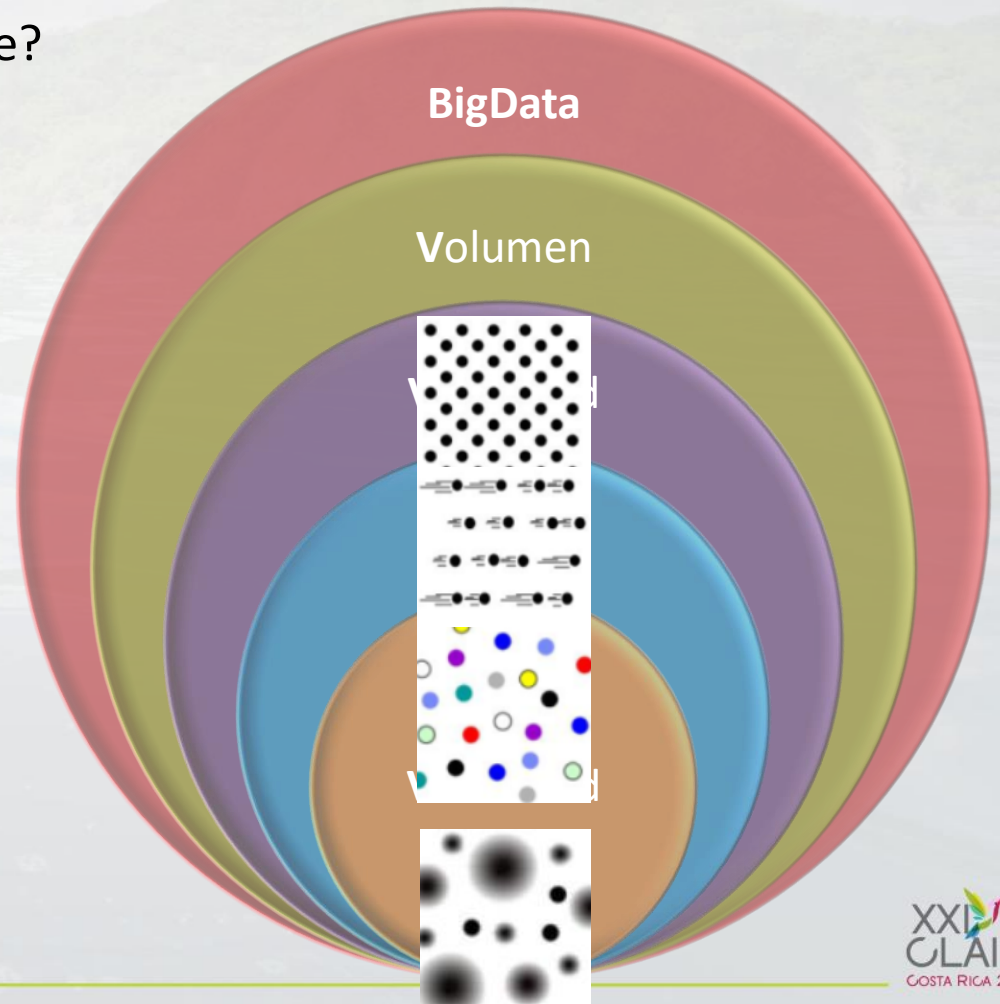
<http://www.gartner.com/it-glossary/big-data/>

Gartner, 2012



# ¿Qué es el BigData?

- ¿BigData ≠ Business Intelligence?
- 2001: 3V (Doug Laney)
- 2012: 4V o 5V...



# BigData: **V**olumen

1 Bit = Binary Digit  
8 Bits = 1 Byte  
1.000 Bytes = 1 Kilobyte  
1.000 Kilobytes = 1 Megabyte  
1.000 Megabytes = 1 Gigabyte  
1.000 Gigabytes = 1 Terabyte  
1.000 Terabytes = 1 Petabyte  
1.000 Petabytes = 1 Exabyte  
1.000 Exabytes = 1 Zettabyte  
1.000 Zettabytes = 1 Yottabyte  
1.000 Yottabytes = 1 Brontobyte  
1.000 Brontobytes = 1 Geopbyte

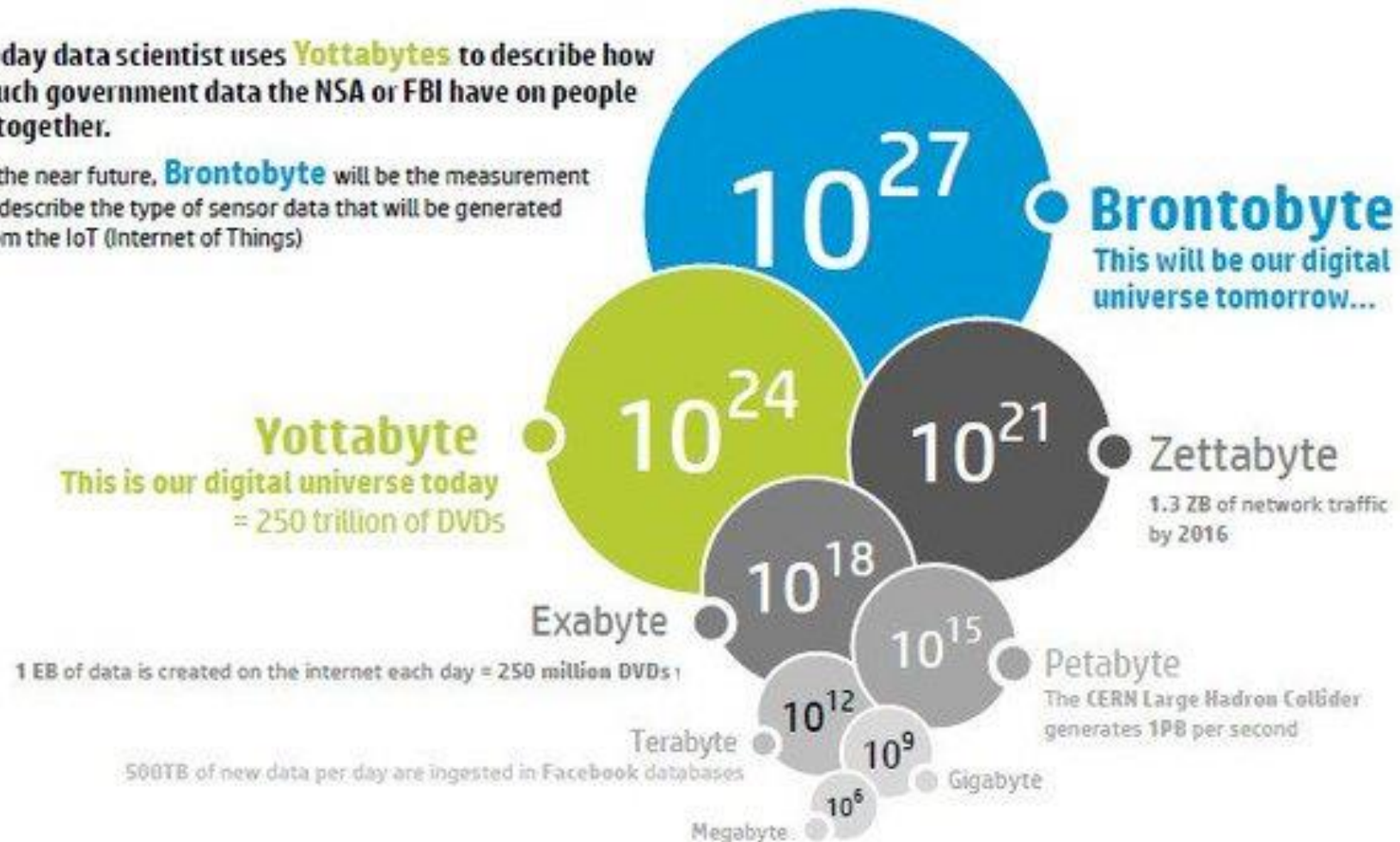




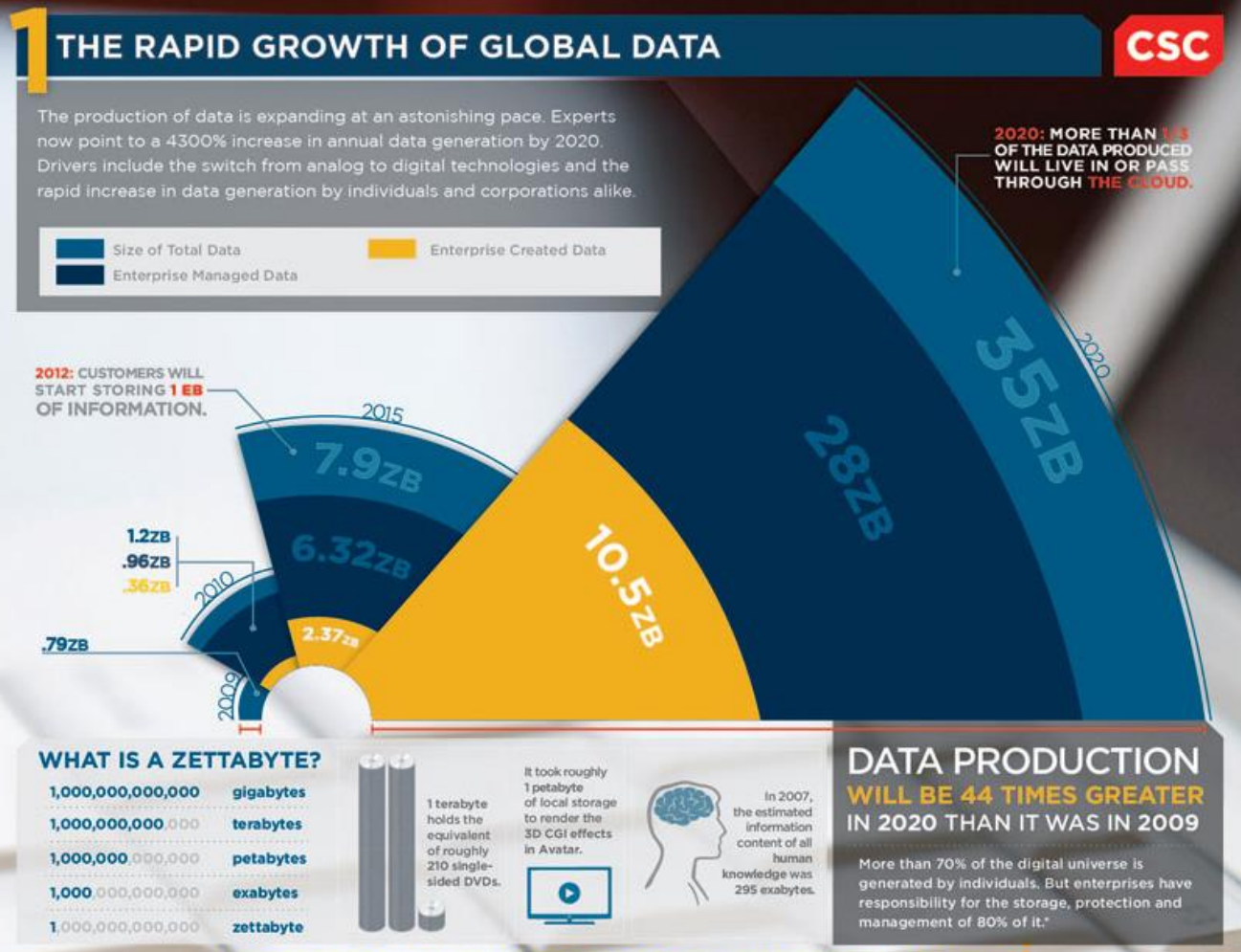
# BigData: **V**olumen

Today data scientist uses **Yottabytes** to describe how much government data the NSA or FBI have on people altogether.

In the near future, **Brontobyte** will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things)



# BigData: Volumen

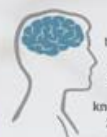


### WHAT IS A ZETTABYTE?

- 1,000,000,000,000 gigabytes
- 1,000,000,000,000 terabytes
- 1,000,000,000,000 petabytes
- 1,000,000,000,000 exabytes
- 1,000,000,000,000 zettabyte

It took roughly 1 terabyte of local storage to render the 3D CGI effects in Avatar.

1 terabyte holds the equivalent of roughly 210 single-sided DVDs.



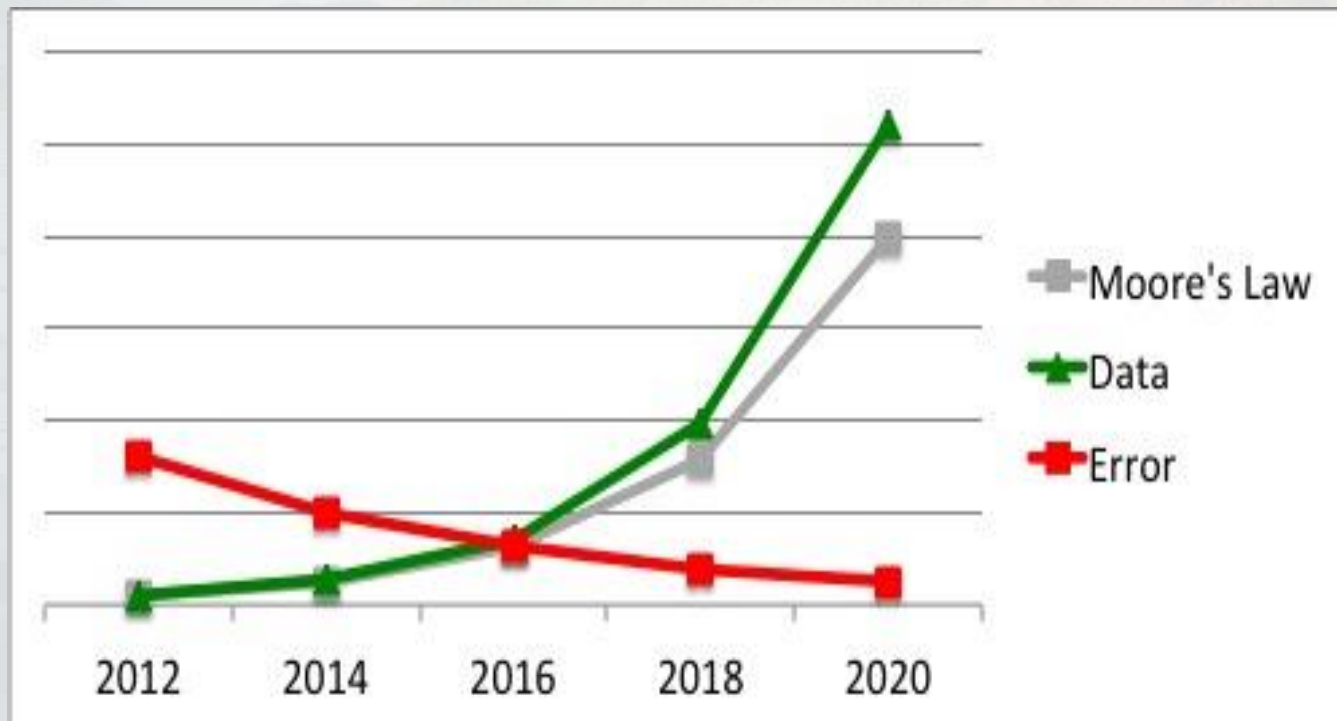
In 2007, the estimated information content of all human knowledge was 295 exabytes.

### DATA PRODUCTION WILL BE 44 TIMES GREATER IN 2020 THAN IT WAS IN 2009

More than 70% of the digital universe is generated by individuals. But enterprises have responsibility for the storage, protection and management of 80% of it.\*



# BigData: Volumen



# BigData: **V**olumen

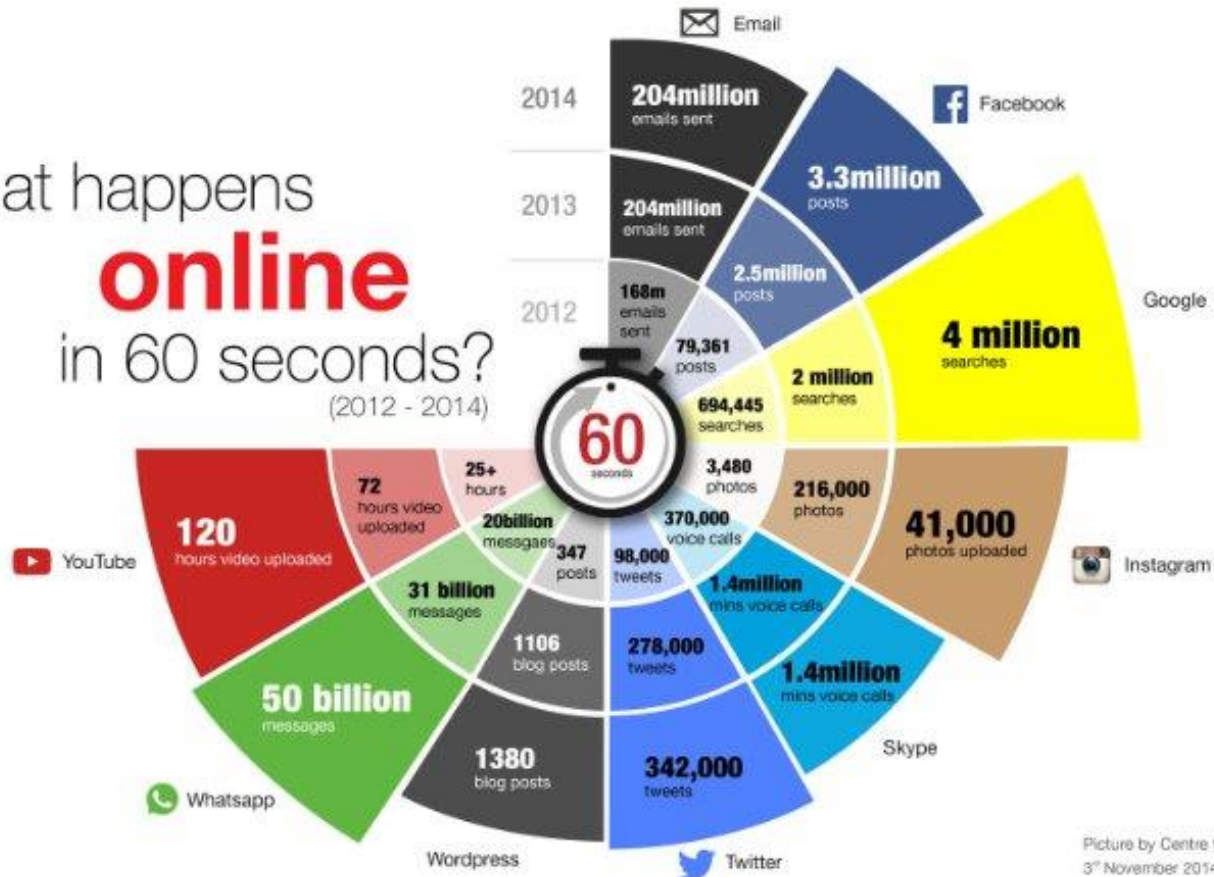


IDC says Digital Universe will be 35 Zettabytes by 2020.

1 Zettabyte = 1,000,000,000,000,000,000 bytes, or 1 billion terrabytes

# BigData: **V**elocidad

What happens  
**online**  
in 60 seconds?  
(2012 - 2014)



Picture by Centre for Learning and Teaching  
3<sup>rd</sup> November 2014



# BigData: **V**ariEDAD



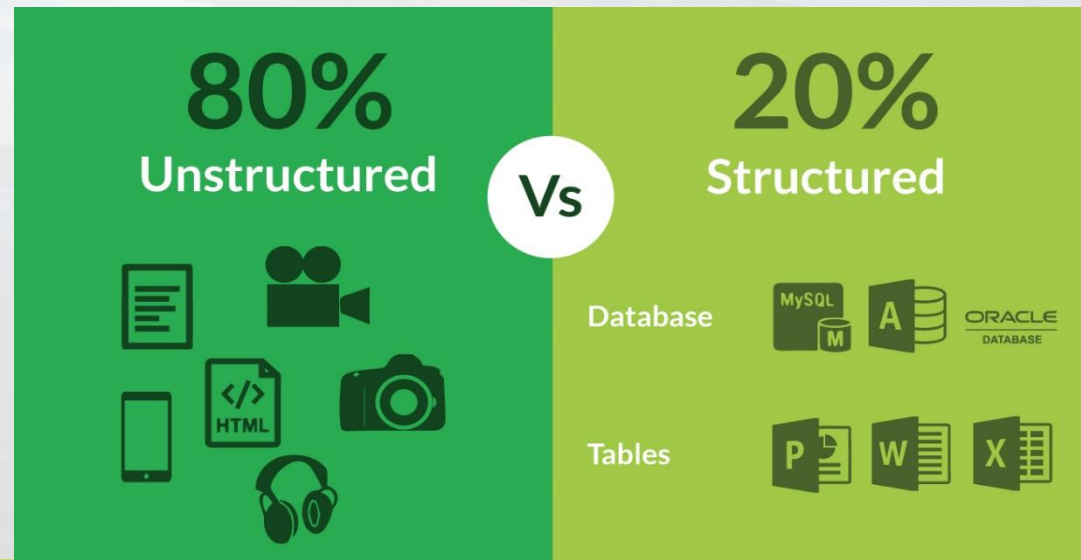
(text, photo, metadata)

- Structured data that fitted into tables & relational databases (e.g. transactional or financial data) relatively straight forward to handle
- Data often unstructured - creates problems for storing, mining & analysing

# BigData: **V**ariedad

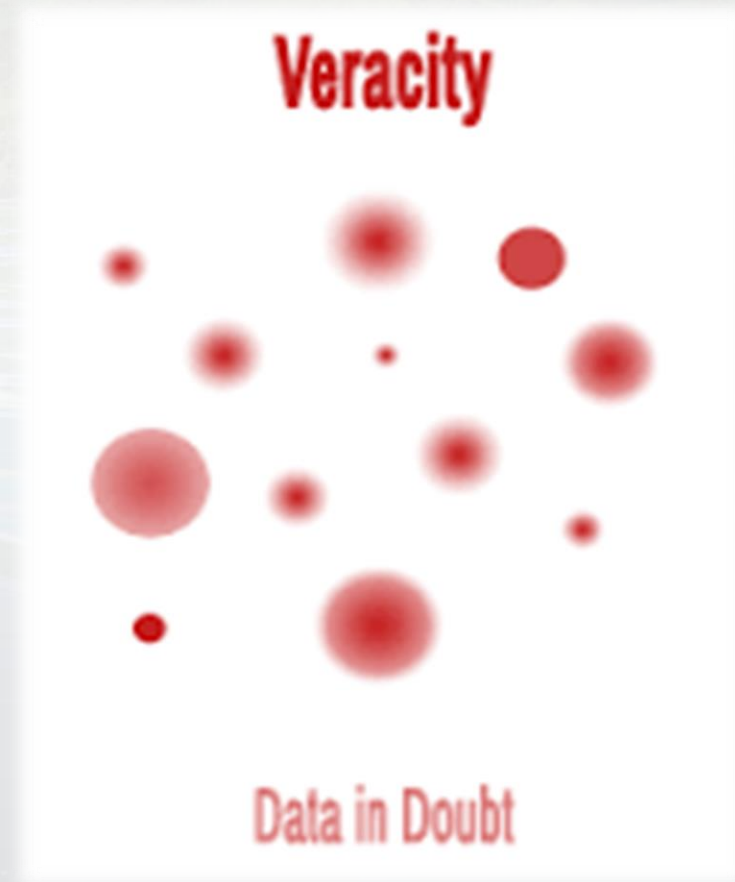
Los datos han pasado de ser estructurados a semi-estructurados o completamente no estructurados.

- **Estructurados:** RDBMS, tablas
- **No estructurados:** Texto, imágenes, vídeo
- **Semi-estructurados:** XML, JSON



# BigData: Veracidad

- **Datos**
  - ... de calidad
  - ... disponibles
  - ... predecibles
- **Evitar**
  - ... ruido
  - ... sesgo
  - ... alteraciones
- **Buscar**
  - ... responsable de los datos
  - ... confianza!





# BigData: Mitos

El Big Data *no* predice el futuro



# BigData: Mitos

El Big Data **no** significa mejores datos



**BIGGER** Data Doesn't  
Mean **BETTER** Data

See why...

# BigData: Mitos

Gran parte de las empresas aún **no han implementado** soluciones de Big Data

*“Big data is like teenage sex;  
everyone talks about it,  
nobody really knows how to do it,  
everyone thinks everyone else is doing it,  
so everyone claims they are doing it”.*

Dan Ariely, Duke University



# Agenda

- Conceptos y mitos del Big Data
- **Impacto del Big Data en distintas industrias**
- Desafíos del Big Data: Roles, infraestructura y regulación
- Casos prácticos de aplicación

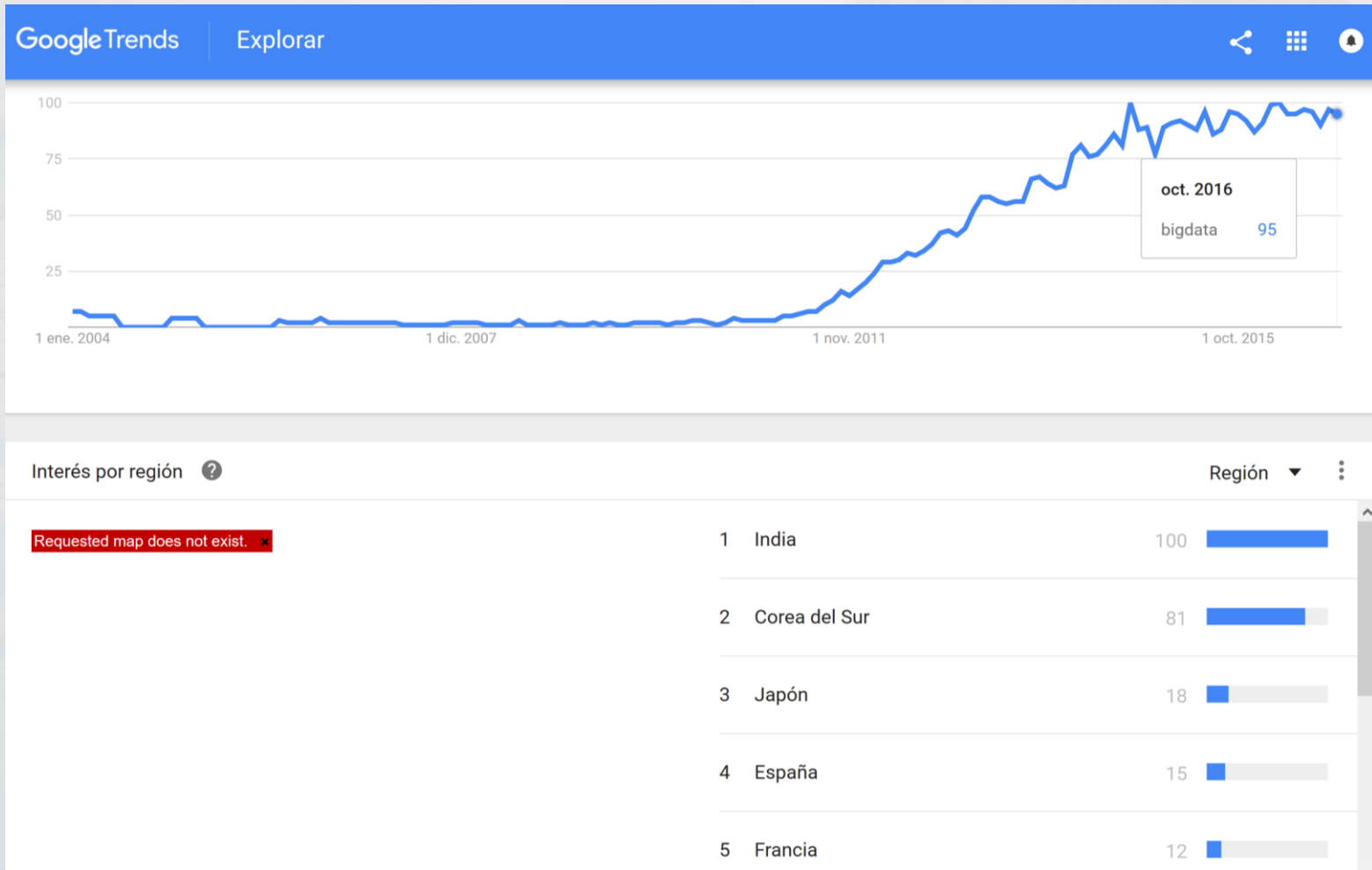


# Impacto del BigData

¿Por qué es interesante el Big Data en las empresas?



# Impacto del BigData





# Impacto del BigData

≡ EL PAÍS 

INTERNACIONAL

EUROPA EE UU MÉXICO AMÉRICA LATINA ORIENTE PRÓXIMO ASIA ÁFRICA FOTOS OPINIÓN BLOGS TITULARES »

## La lealtad del pueblo chino la mide el 'big data'

El Gobierno de Xi Jinping usará datos personales para catalogar a los ciudadanos



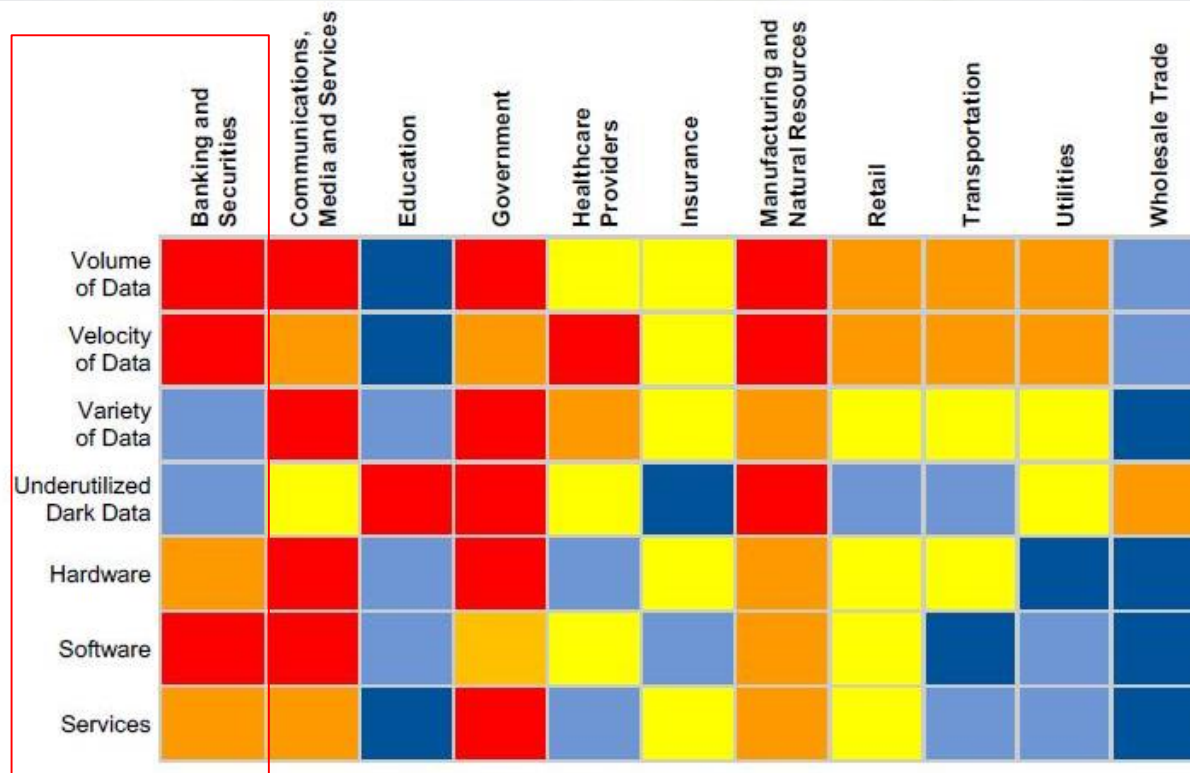
MACARENA VIDAL LIY

Pekín - 23 OCT 2016 - 07:28 CEST

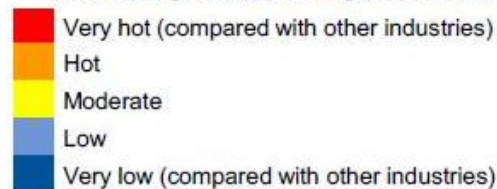
China prepara un sistema disciplinario basado en la recogida exhaustiva de datos personales de personas y de empresas, que se usarán para puntuar la fiabilidad social y la lealtad política de cada individuo y de cada compañía. Ser catalogado como mal ciudadano acarreará castigos tan diversos como la prohibición de alojarse en determinados hoteles, de viajar de la manera más cómoda o de que sus hijos puedan asistir a las mejores escuelas. Un sistema orwelliano que las autoridades chinas quieren tener en marcha para 2020 o incluso antes.

# Impacto del BigData

Oportunidades por industrias...



Potential big data opportunity on each dimension is:



Fuente: Gartner 2012

# Impacto del BigData

Oportunidades por industrias...

	Manu & N. Res.	Media/ Comm	Svcs	Gov.	Edu	Retail	Banking	Insur- ance	Health- care	Trans- portation	Utilities
Enhanced customer experience	52%	78%	66%	43%	76%	83%	77%	77%	73%	69%	44%
Process efficiency	45%	33%	35%	49%	65%	43%	41%	50%	73%	69%	78%
More targeted marketing	43%	89%	53%	17%	41%	78%	66%	58%	-	38%	17%
Cost reduction	42%	33%	35%	37%	35%	30%	41%	31%	45%	56%	61%
Improved risk management	14%	22%	29%	29%	35%	22%	52%	58%	55%	31%	61%
New products	23%	67%	37%	14%	24%	35%	27%	50%	-	19%	33%
Developing information products	26%	33%	44%	31%	12%	22%	23%	19%	9%	19%	11%
Enhanced security capabilities	17%	22%	21%	34%	29%	13%	27%	27%	9%	19%	28%
Regulatory compliance	11%	22%	18%	23%	18%	9%	25%	23%	27%	31%	44%
<i>n</i> =	65	9	62	35	17	23	44	26	11	16	18

Fuente: Gartner 2015



# Impacto del BigData

Oportunidades en las áreas de **riesgo de crédito** de los bancos



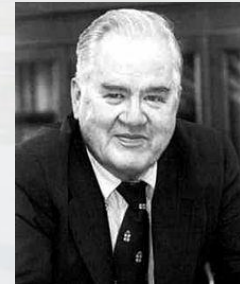
# Agenda

- Conceptos y mitos del Big Data
- Impacto del Big Data en distintas industrias
- **Desafíos del Big Data: roles y regulación**
- Casos prácticos de aplicación



# Desafíos del BigData

*“La combinación de algunos datos y un **deseo doloroso** de una respuesta, no asegura que una respuesta razonable pueda ser extraída de una determinada masa de datos”.*

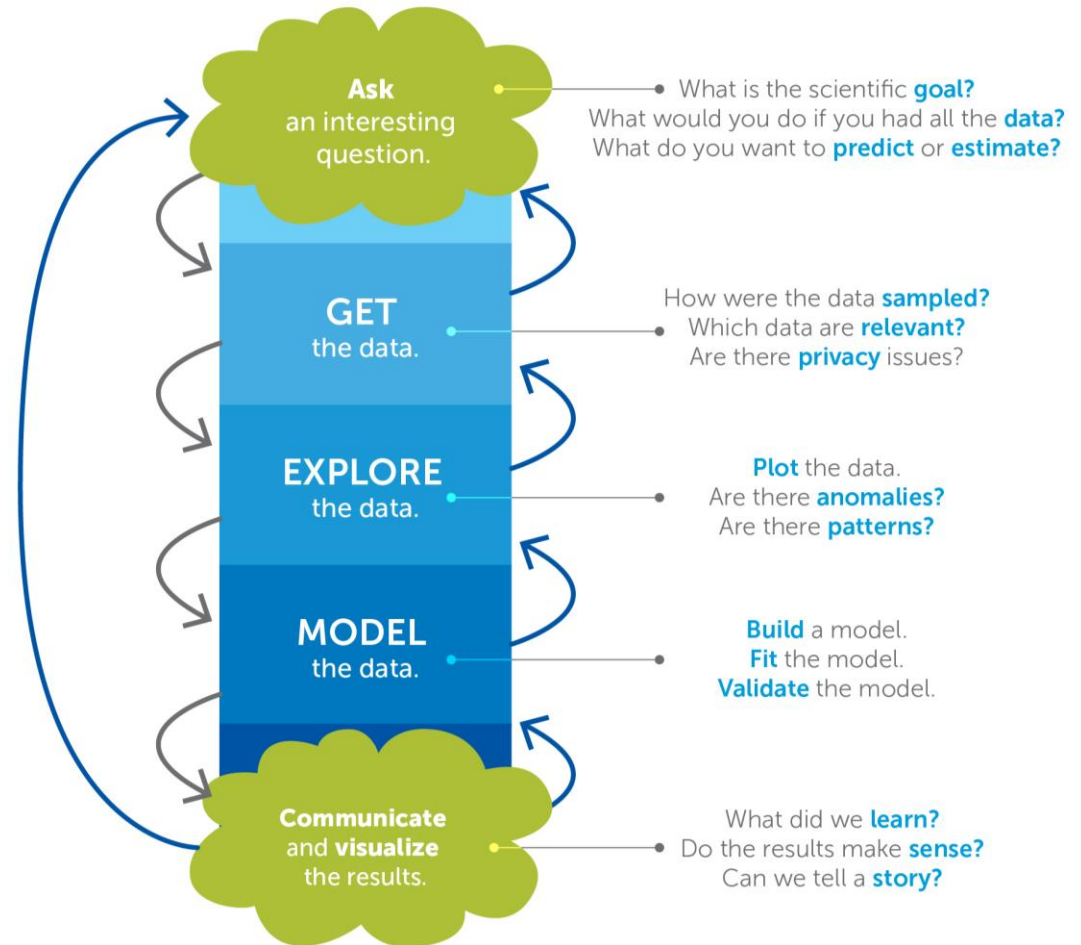


John Turkey  
(1915-2000)



# Desafíos del BigData

## The Data Science Process



Derived from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course <http://cs109.org/>.

# Desafíos del BigData

HELPING YOU REACH  
VELOCITY

## Data Scientist

Data Science allows front offices to better predict what and when consumers are likely to buy. The ability to write algorithms that find relationships in datasets is usable to provide actionable insight.



### Urgent Need

Data Scientists - those with the technical savvy and analytical chops to derive meaning from all the information- are in high demand

### Skills by the Numbers

The skills and talents that make a fantastic Data Scientist

Complex Formulas	40%
Consumer Psychology	25%
Business Acumen	25%
Programming Languages	10%

### The Challenge

- Data Mining
- Analysis
- Communication

### Industry Niche Titles

- Financial Institutions/ Decision Scientist
- Retailers/Omni Channel Expert
- Marketing Agencies/Consumer Behaviour Analyst
- Ecommerce/Analytics Expert

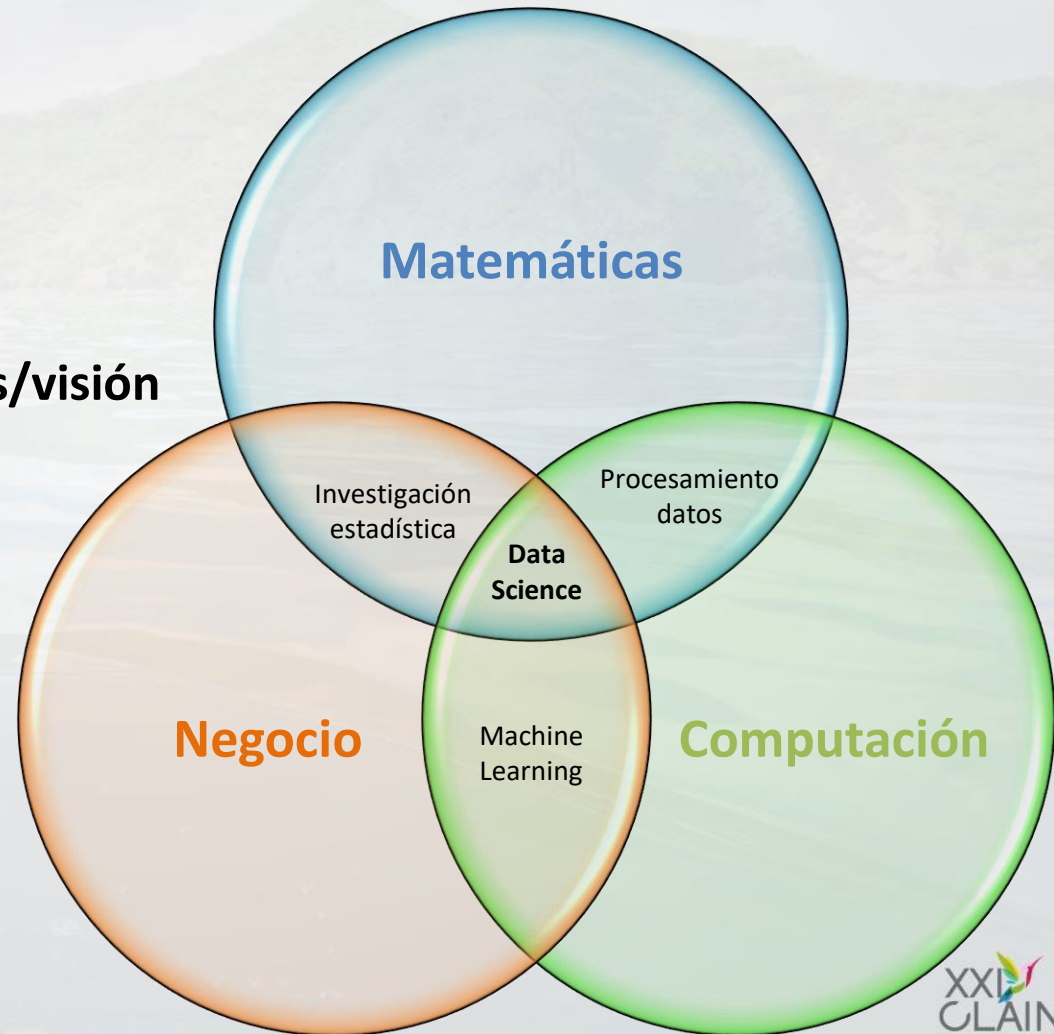
### Did you Know?

Google's Eric Schmidt claims that every two days now we create as much information as we did from the dawn of civilization up until 2003



# Desafíos del BigData: Roles

- **Sólida formación**
  - ... matemáticas
  - ... estadísticas
  - ... programación
- **Comprender el valor de datos/visión**
  - ... curiosidad
  - ... software
  - ... analytcis
- **Métricas y negocio**
  - ... visión cliente
  - ... comunicación
  - ... ética
  - ... *think outside the box*





# Desafíos del BigData: Roles

Algunos ejemplos de relacionamiento y trabajo compartido (*lista no taxativa*)

## CRO

### Riesgo

- Riesgo de Mercado, Crédito y Operacional
- Pruebas regulatorias de estrés
- Fraudes y detección y prevención de anomalías

## CFO

### Finanzas

- Rentabilidad
- Riesgo ajustado al presupuesto planificado y proyecciones
- Analytics para el control, eficiencia y reducción del capital regulatorio
- Reportes Regulatorios



## CMO

### Clientes

- Gestión de los precios
- Suscripción y Publicidad
- Análisis de los "sentimientos" de los Clientes
- Abandono y retención de Clientes
- Marketing

## COO

### Operaciones

- Operaciones de Negocio
- Racionalización – personas y procesos
- Eficiencia Operacional

## CHRO

### Colaboradores

- Retención
- Planificación de Dotación
- Seguridad
- Reclutamiento

# Desafíos del BigData: Regulación

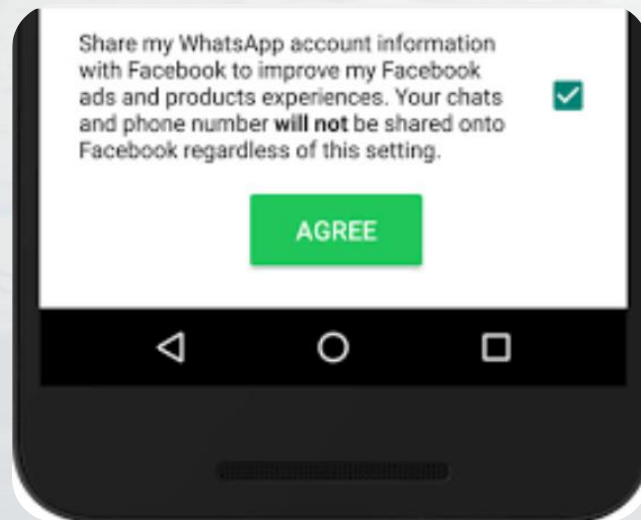
La European Data Protection Supervisor (EDPS) cree que el desarrollo **responsable y sostenible** de grandes volúmenes de datos debe confiar en 4 elementos esenciales:

- 1) Las organizaciones deberán ser mucho más **transparentes** acerca de cómo procesan los datos personales.
- 2) Se deberá permitir a los usuarios un mayor grado de **control** sobre cómo se utiliza su información.
- 3) Se **diseñará** una política de protección de datos de los usuarios en sus productos y servicios.
- 4) Las compañías deberán **responsabilizarse** más de todos los procesos que lleven a cabo.



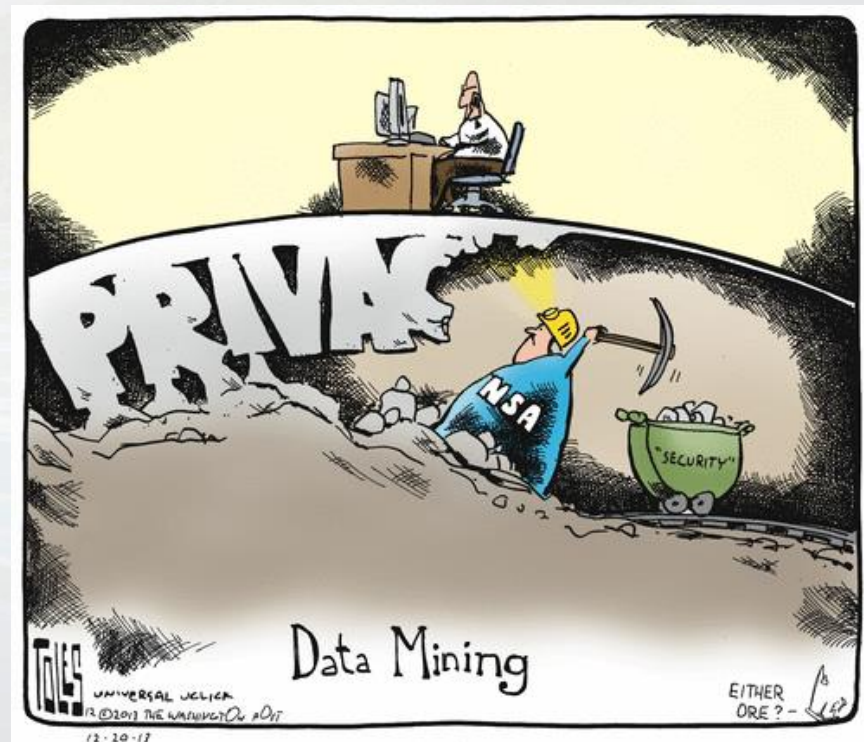
# Desafíos del BigData: Regulación

- Legislación ambigua: “En tierra de nadie” ...
- Principio de “uso de datos para el propósito que se ha contratado”
- Sólo un **61%** de aplicaciones móviles muestran políticas de privacidad
- Protección de privacidad con métodos de **notificación y consentimiento**



# Desafíos del BigData: Regulación

- El principal responsable es el propio **usuario**...
- Generalmente el **regulador bancario** se focaliza en los modelos de riesgo en originación y provisiones.  
... Oportunidades en **modelos de gestión**: seguimiento, alertas y cobranzas...





# Agenda

- Conceptos y mitos del Big Data
- Impacto del Big Data en distintas industrias
- Desafíos del Big Data: roles y regulación
- **Casos prácticos de aplicación**



# Casos prácticos de aplicación

- Existen mucho mitos sobre **la aplicación del Big Data** en los bancos
- Parte de estos mitos está asociado a la multiplicación de las **Fintech**
- Muchas Fintech se han desarrollado para **cubrir espacios** que la banca tradicional no cubre o no atiende de **forma ágil**, aprovechando el valor de los datos e innovando (start-up)



- Salvo algunas excepciones de éxito, muchas de Fintech tienen el objetivo de “venderse” o ser **absorbidas por bancos**



# Casos prácticos de aplicación

CaixaBank: Proyecto de pre-aprobación para clientes y no clientes (AIS 2016)



# Casos prácticos de aplicación

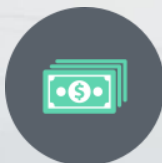
**CaixaBank:** Proyecto de pre-aprobación para clientes y no clientes (AIS 2016)



Para los no clientes, se pregunta, entre otros, por la dirección postal de la vivienda habitual



Se realiza una valoración del valor de la vivienda.



El valor de la vivienda es un gran indicador del nivel de ingresos del solicitante

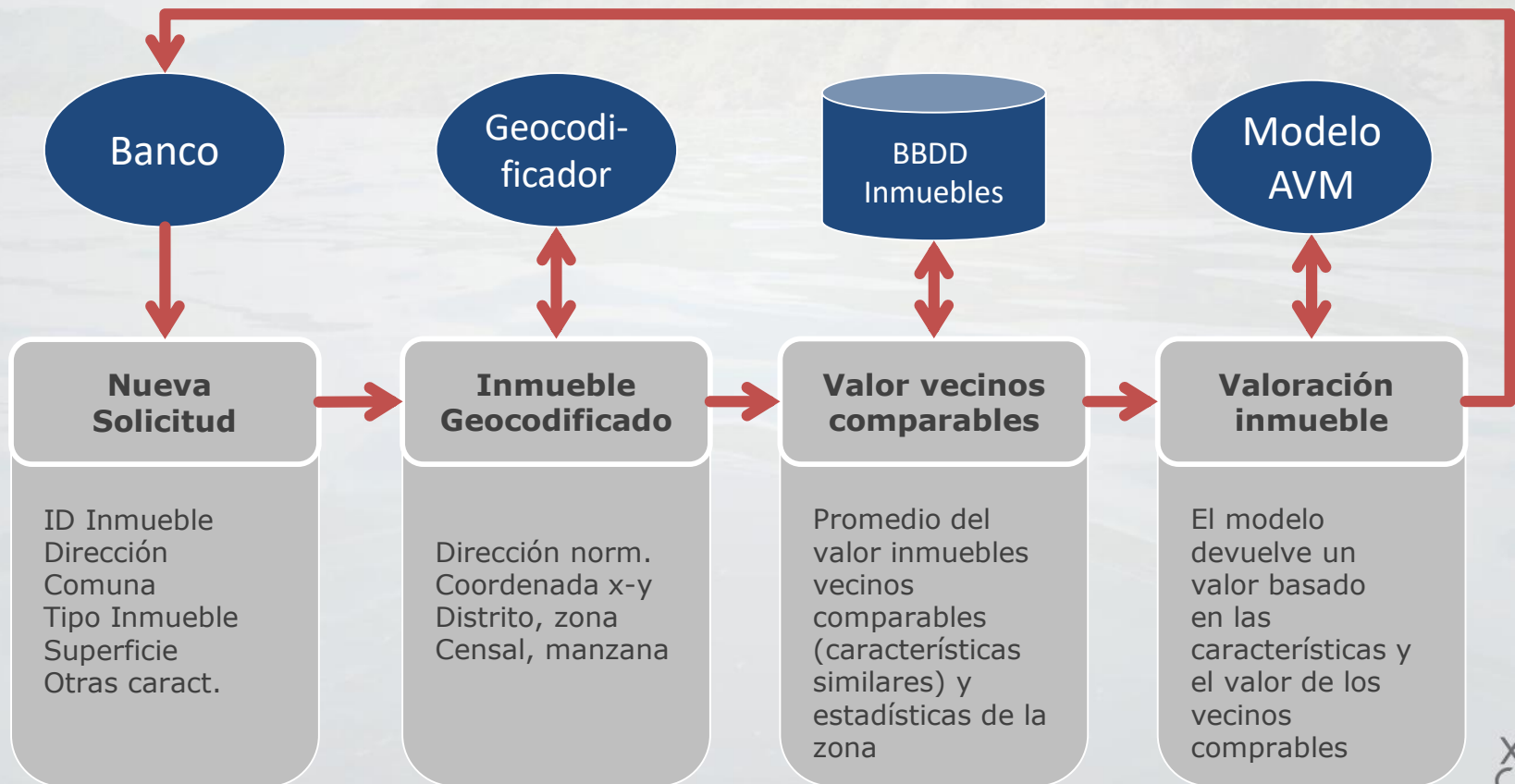


Con este ingreso estimado, se determina el importe máximo a conceder



# Casos prácticos de aplicación

**Solvía, Banco Sabadell:** Proyecto Valoración Automática Inmuebles (AIS 2017)



# Casos prácticos de aplicación

**Solvía, Banco Sabadell:** Proyecto Valoración Automática Inmuebles (AIS 2017)



- En casas, el error individual máximo:  $\pm 5\%$
- En departamentos, el error individual máximo:  $\pm 7\%$

# Casos prácticos de aplicación

## CaixaBank: Proyecto de alertas tempranas (AIS 2017)



- El deterioro de un cliente puede comenzar a constatarse de forma **poco generalizada**. Es decir, con el empeoramiento de sólo un par de características a la vez dentro de una basta batería de alertas.
- Con las regresiones logísticas clásicas, **variables poco significativas** quedan fuera del algoritmo. Por tanto, el modelo sería ciego a este tenue empeoramiento.
- Existen técnicas para aumentar el número de variables dentro de un modelo, como las regresiones sobre *subscores*, pero **no son óptimas...**

# Casos prácticos de aplicación

## CaixaBank: Proyecto de alertas tempranas (AIS 2017)



- Aprovechar **toda la información disponible** en cuanto a la detonación de alertas, seleccionando una muestra histórica.
- El sistema de alertas a diseñar debe permitir **adelantarse en todos los ámbitos** al evento que se quiere prevenir. Claves:
  - ✓ Definir adecuadamente el horizonte de observación del evento. Ni muy largo ni muy corto y alineado con la regulación.
  - ✓ Adaptarse a la segmentación del banco.
  - ✓ Definir el evento a predecir acorde al nivel de deterioro del cliente
- Desarrollo de modelos de **Machine Learning**.



# Casos prácticos de aplicación

## CaixaBank: Proyecto de alertas tempranas (AIS 2017)



- Modelos muy discriminantes (Gini 90%)
- Modelos con un set de aprox. 50 variables.

# Casos prácticos de aplicación

- **Kutxabank:** Proyecto AIS KnowU (En fase conceptual...)

Solución que permite **conocer la situación financiera** de no clientes o clientes poco vinculados, aprovechando la información transaccional de los bancos con los cuales trabaja el cliente (*scraping*).



# Información de Contacto

## **Leandro Fernández Lacalle**

Gerente General  
AIS Chile y Región Andina  
[leandro.fernandez@ais-int.com](mailto:leandro.fernandez@ais-int.com)



[www.ais-int.com](http://www.ais-int.com)



[@GrupoAIS](https://twitter.com/GrupoAIS)



<http://www.linkedin.com/company/ais>



[Riesgo de Crédito \(grupo\)](#)

# Información de Contacto



Empresa líder y pionera en sistemas de apoyo a la toma de decisiones



Fundada en 1987 en Barcelona



Más de 200 clientes en 20 países



Oficinas regionales en Chile, Argentina, Colombia, México y Portugal



Socio colaborador del Big Data Center of Excellence de Barcelona





# Preguntas y Respuestas