

www.bsc.es



**Barcelona  
Supercomputing  
Center**  
Centro Nacional de Supercomputación

# La tecnología detrás del *Big Data*

**Mario Macías, PhD**

Autonomic Systems and e-Business Platforms

<http://www.xaas.guru>

Septiembre 2014



“El Big Data es como el sexo para los adolescentes:  
todos hablan de ello,  
nadie sabe del todo cómo funciona,  
todos piensan que los demás lo hacen  
→ todos dicen practicarlo”

*Alguien en internet*



La tecnología detrás del Big Data. Mario Macías. [www.xaas.guru](http://www.xaas.guru)

## De dónde venimos

**Integridad Referencial**

**Autor:** R  
**Título:** E  
**Editorial:** A  
**Año:** 2  
**Nº de páginas:** 9  
**Encuadernación:** T  
**ISBN:** 9  
**Lengua:** C

**Localización:** Filosofía 3.14 (16)

**Referencia**

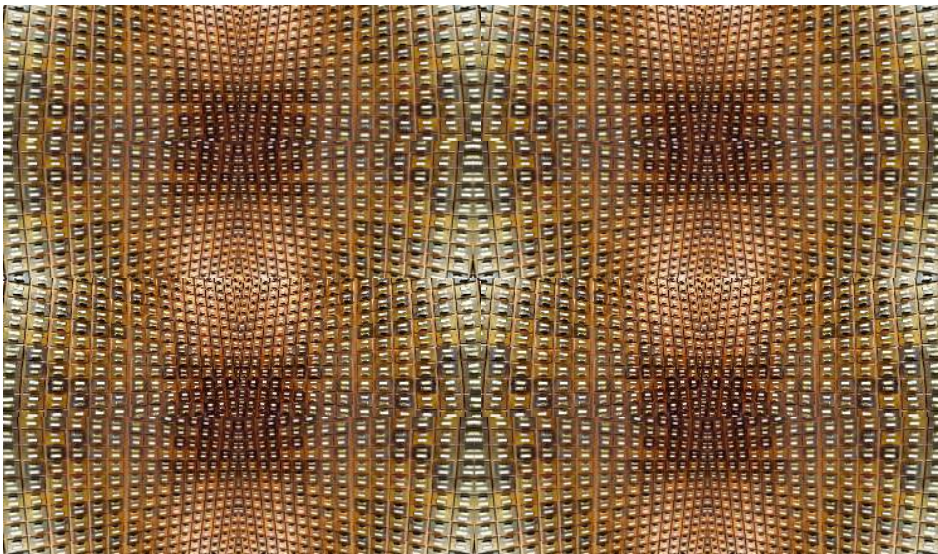


La tecnología detrás del Big Data. Mario Macías. [www.xaas.guru](http://www.xaas.guru)

Si esto es una base de datos...

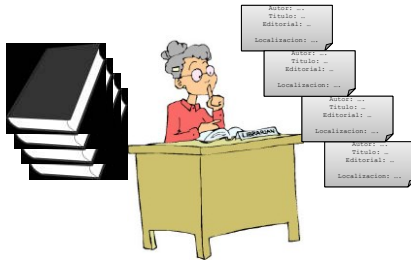


...¿es esto “Big Data”?

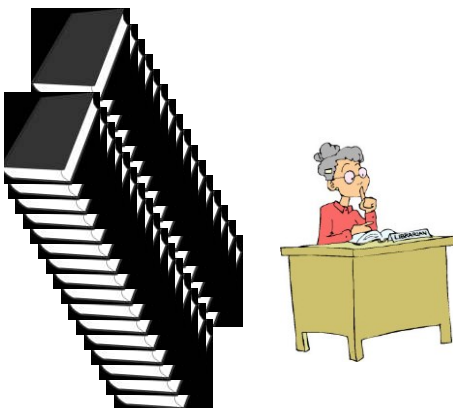




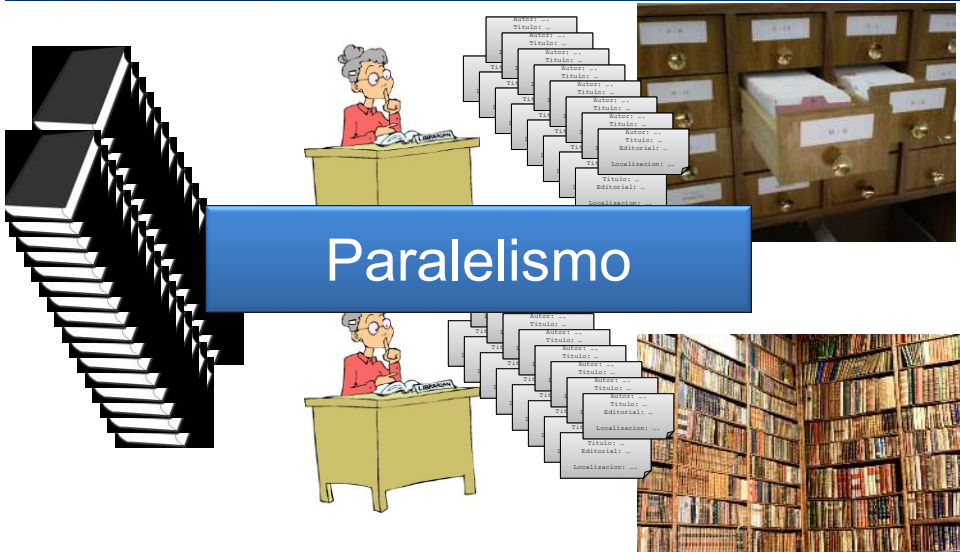
## Trabajando con datos relacionales



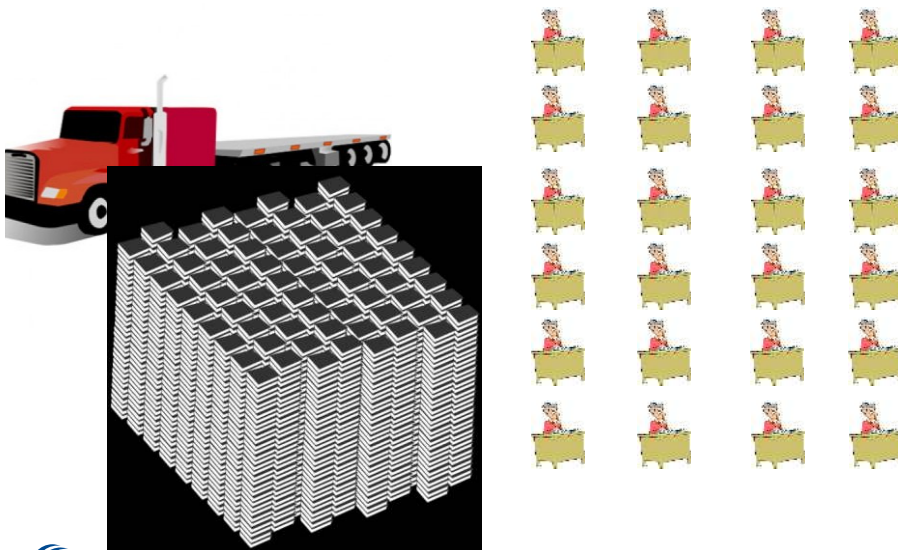
## Trabajando con MUCHOS datos relacionales



## Trabajando con MUCHOS datos relacionales



## Trabajando con DEMASIADOS datos relacionales



Trabajando con DEMASIADOS datos relacionales



“Tenemos un problema  
de Big Data cuando el  
tamaño de los datos es  
un problema”

## Antes: almacén de datos



## Ahora: ~~caja de arena~~ vertedero de datos

- Redes sociales: fotos, comentarios, “me gusta”, #hashtags...
- Agregadores de noticias, agencias de información, “Open Data”...
- Ciudades sensorizadas: “Smart cities”
- Dispositivos móviles

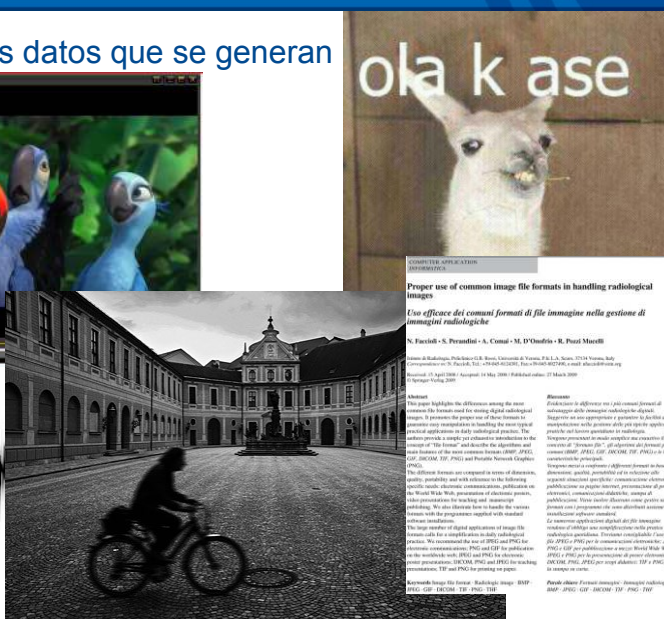
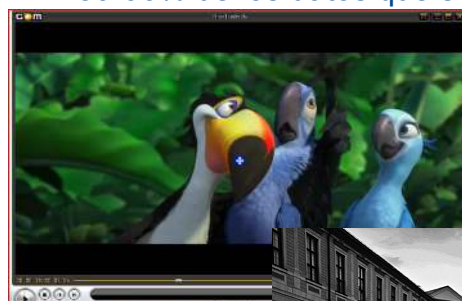
*(y un largo etcétera...)*

## Datos estructurados

**Autor:** René Descartes  
**Título:** El Discurso del Método  
**Editorial:** AKAL  
**Año:** 2007  
**Nº de páginas:** 96 págs.  
**Encuadernación:** Tapa blanda  
**ISBN:** 9788446027898  
**Lengua:** Castellano  
  
**Localización:** Filosofía 3.14 (16)

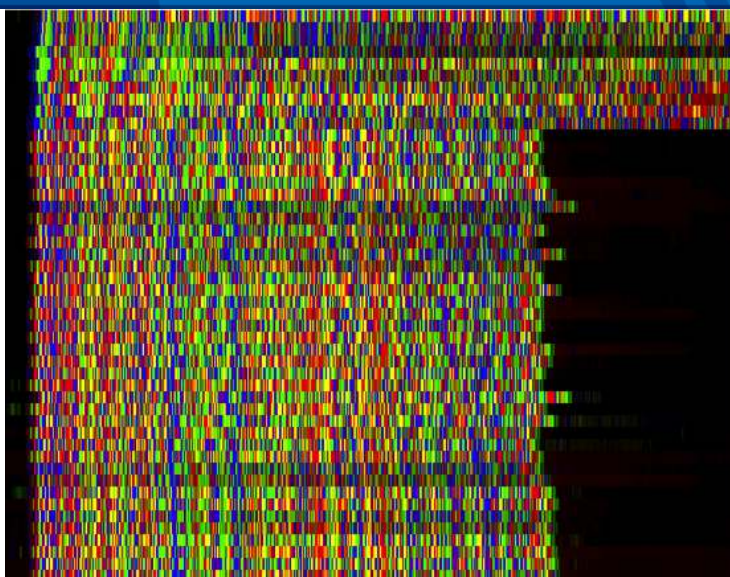
## Datos desestructurados

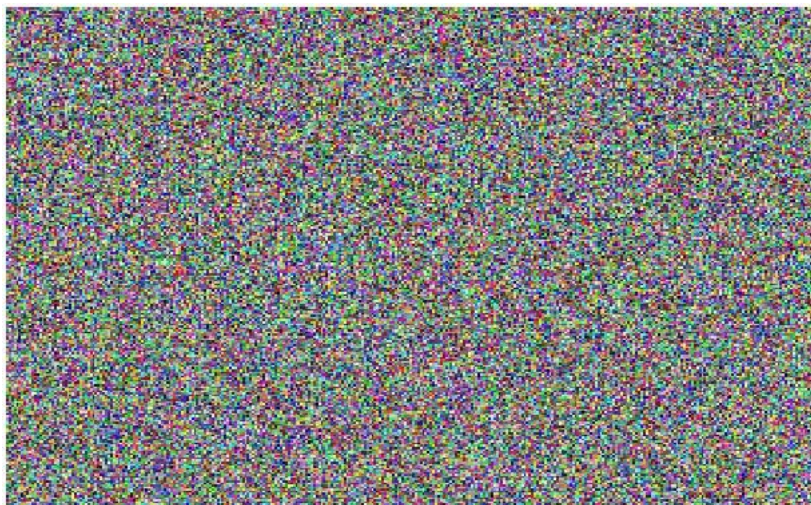
“ El 80-90% de los datos que se generan



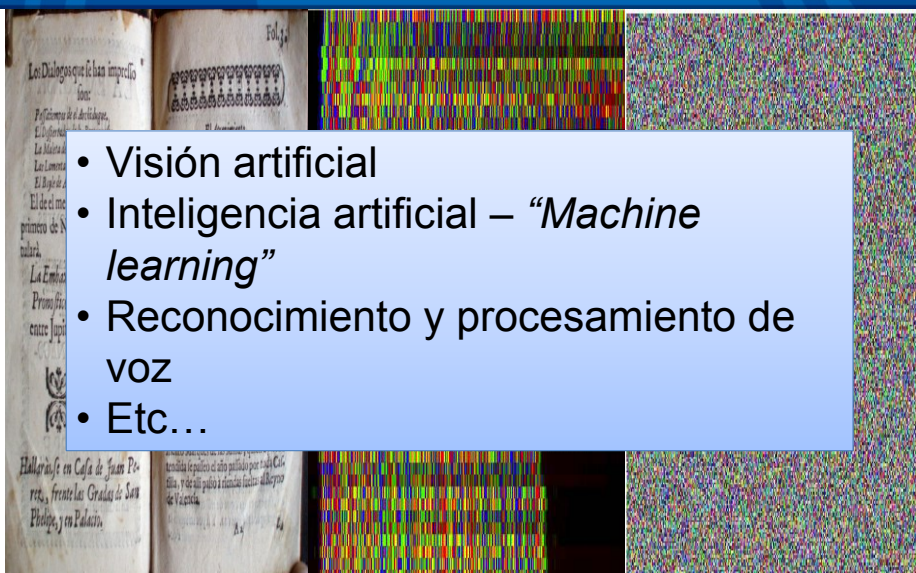


## ¿Cómo sacar información de los datos desestructurados?





## Los datos solo son datos



# Big Data =

- «Muchos datos: “Big”
- «Limitaciones de tiempo
- «Límites de escalabilidad
- «Diversas fuentes
- «Desestructuración
- «Multidisciplinaridad



La tecnología detrás del Big Data. Mario Macías. [www.xaas.guru](http://www.xaas.guru)

[www.bsc.es](http://www.bsc.es)



**Barcelona  
Supercomputing  
Center**  
Centro Nacional de Supercomputación

## Las tecnologías detrás del Big Data: algunas pinceladas



## Una pila de software para el Big Data

### Sistema de ficheros distribuido



La tecnología detrás del Big Data. Mario Macías. [www.xaas.guru](http://www.xaas.guru)

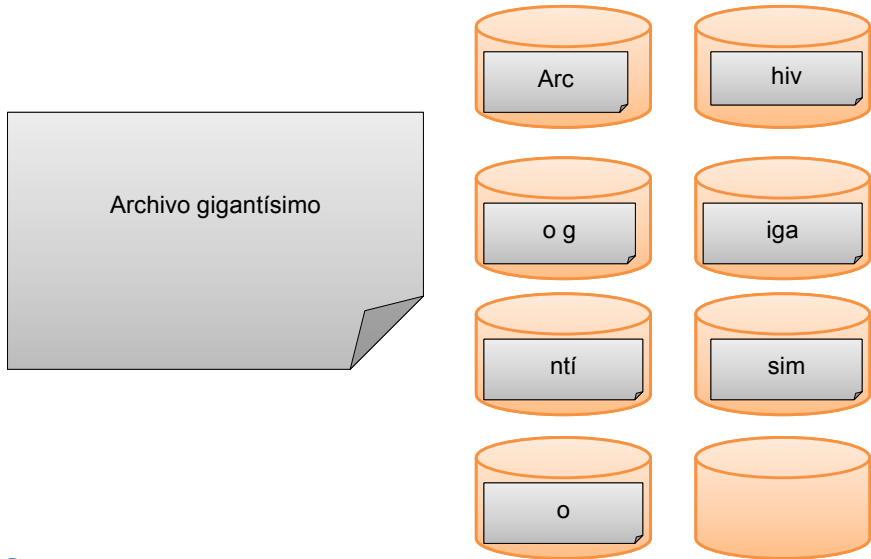
## Sistema de ficheros

- “ Un sistema de ficheros organiza cómo los datos están guardados en una unidad de almacenamiento



La tecnología detrás del Big Data. Mario Macías. [www.xaas.guru](http://www.xaas.guru)





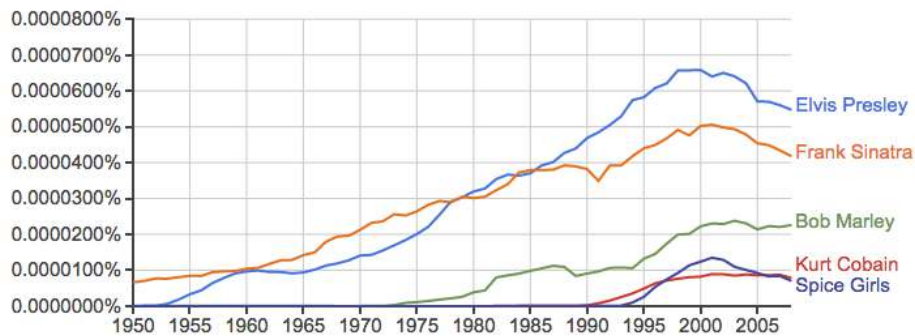
## Una pila de software para el Big Data





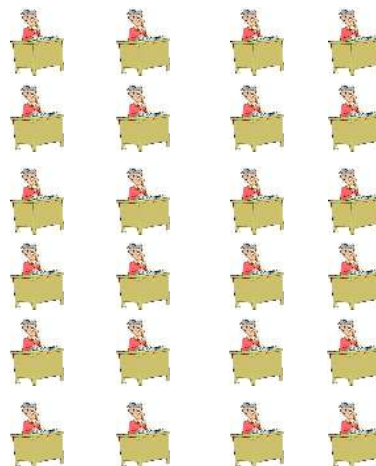
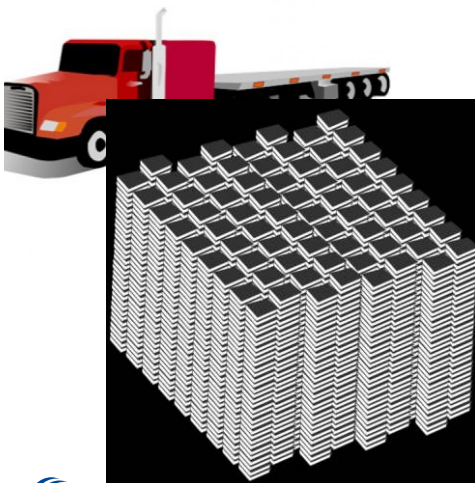
## Google books Ngram Viewer

Graph these comma-separated phrases:  ☐ case-  
between  and  from the corpus  with smoothing of



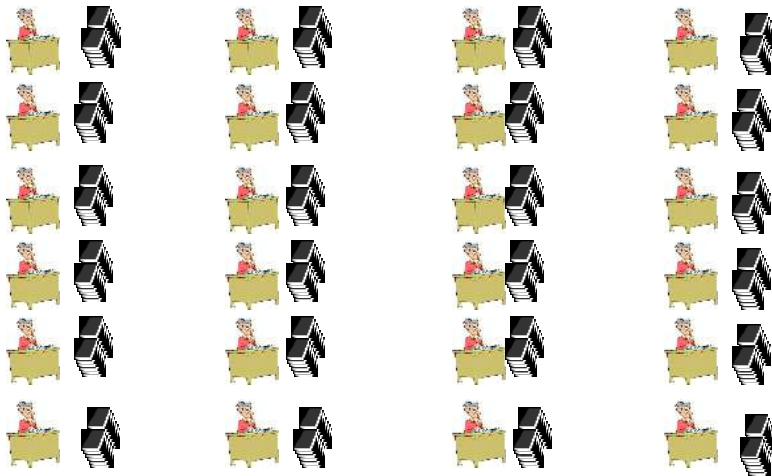
La tecnología detrás del Big Data. Mario Macías. [www.xaas.guru](http://www.xaas.guru)

¿Cuántas veces se repite la palabra “pelota” según el año?



La tecnología detrás del Big Data. Mario Macías. [www.xaas.guru](http://www.xaas.guru)

## Paso 1. Dividir el trabajo



## Paso 2. Map



Año Libro	Apariciones
1920	15
1950	22
1955	33
1958	32
1960	40



Año Libro	Apariciones
1923	12
1950	20
1954	30
1958	29
1959	35

Decenas  
de tablas

Algunas  
contienen  
datos  
sobre los  
mismos  
años

### Paso 3. Reduce

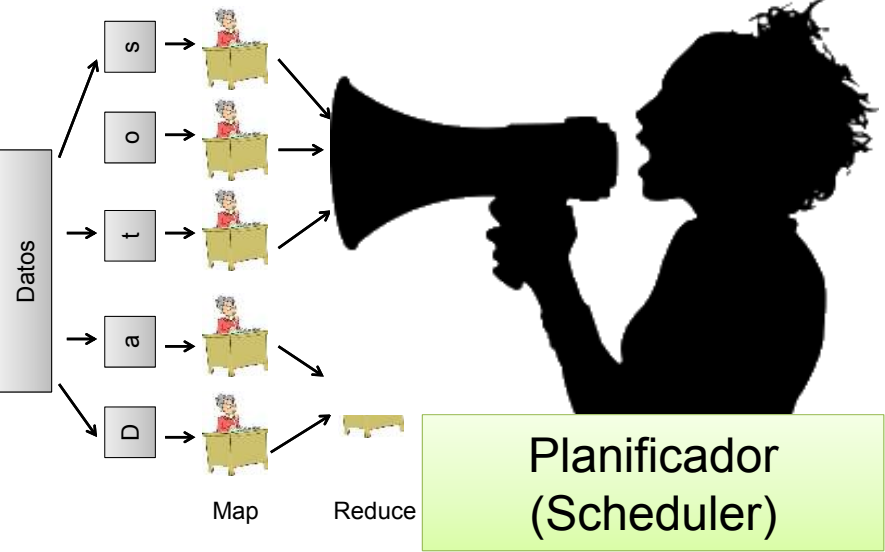
Año Libro	Apariciones
1920	15
1950	22
1955	33
1958	32
1960	40

Año Libro	Apariciones
1923	12
1950	20
1954	30
1958	29
1959	35



Año Libro	Apariciones
1920	15
1950	42
1954	30
1955	33
1958	61
1959	35
1960	40

### Map/Reduce. Visión general





## Una pila de software para el Big Data

Capa de abstracción

Modelo de programación

Sistema de ficheros distribuido



La tecnología detrás del Big Data. Mario Macías. [www.xaas.guru](http://www.xaas.guru)

Mario Macías Lloret

## Del bit a la Nube



2,99€

amazonkindle

Google play

BARNES  
& NOBLE



iBooks

[www.bsc.es](http://www.bsc.es)



**Barcelona  
Supercomputing  
Center**  
*Centro Nacional de Supercomputación*

# ¡Gracias por tu atención!

**Mario Macías, PhD**  
Autonomic Systems and e-Business Platforms  
Junio 2014