

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

De dónde venimos



Integridad Referencial

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

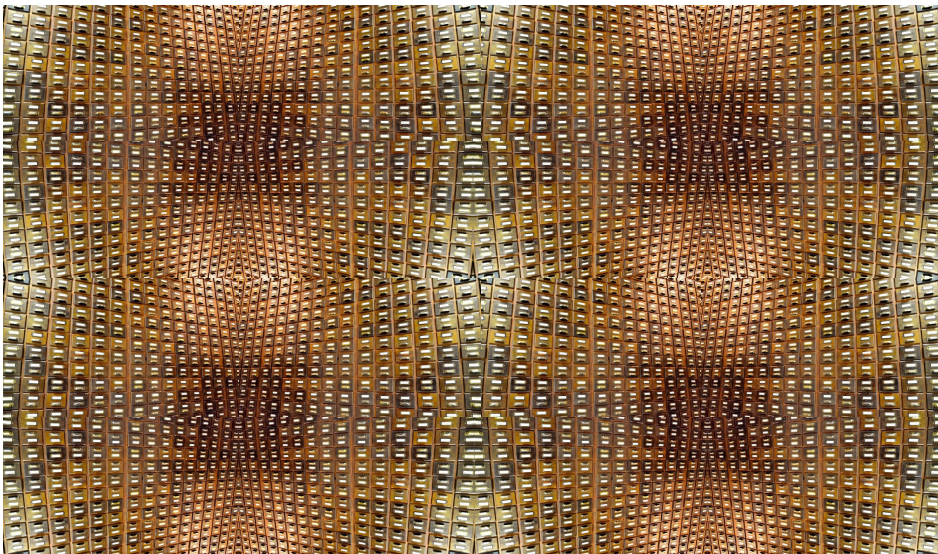
Localización: Filosofía 3.14 (16)

Referencia

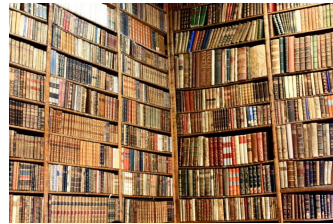
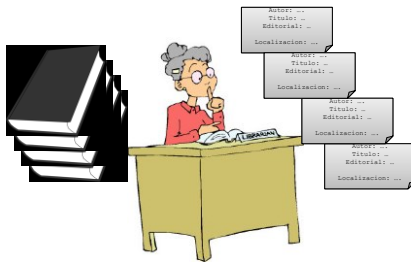
Si esto es una base de datos...



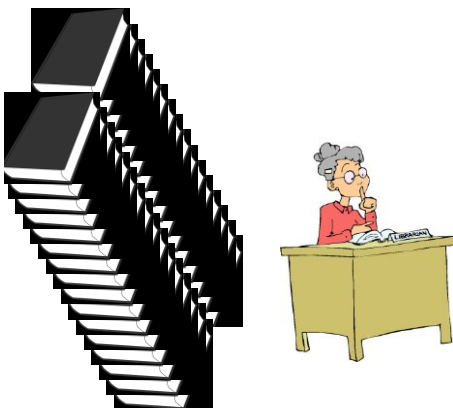
...¿es esto “Big Data”?



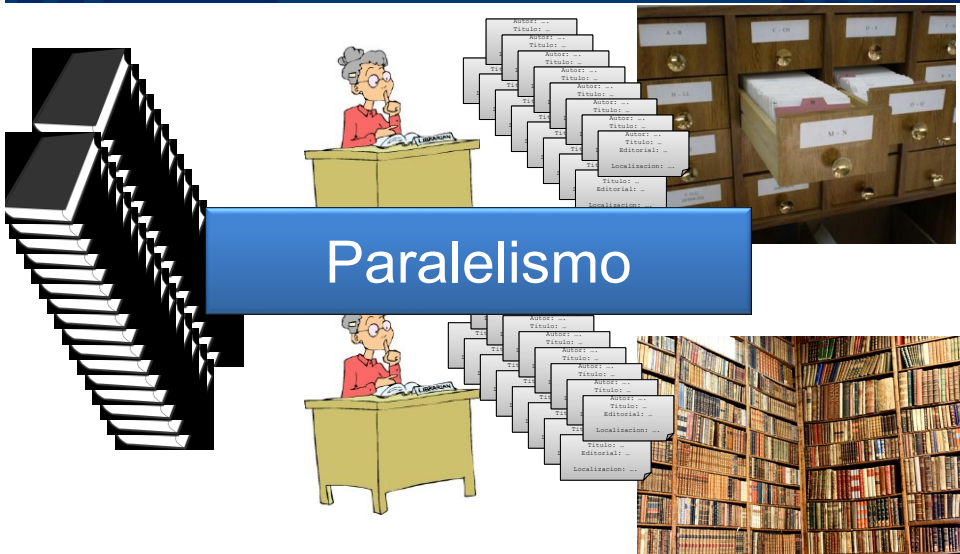
Trabajando con datos relacionales



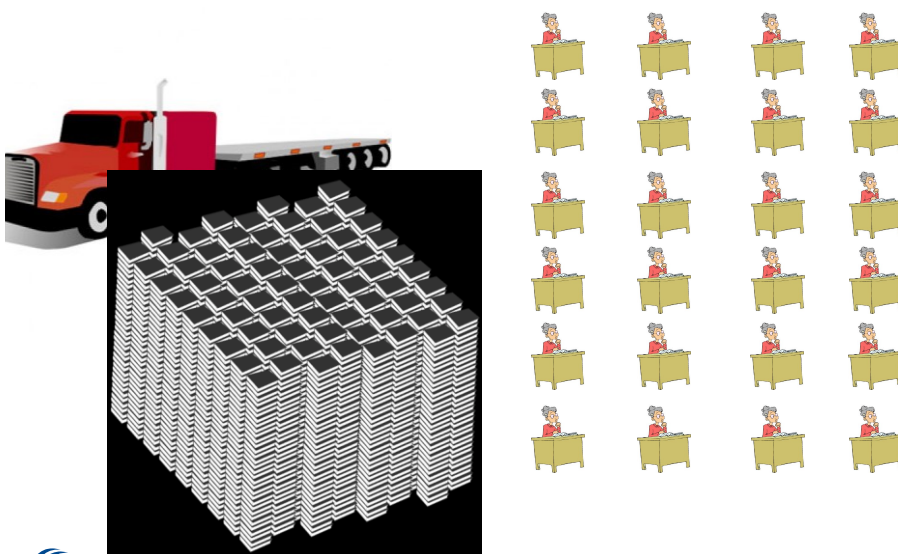
Trabajando con MUCHOS datos relacionales



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

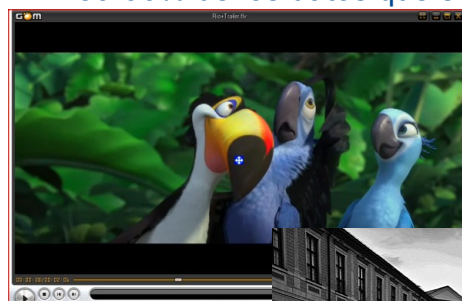
(y un largo etcétera...)

Datos estructurados

Autor: René Descartes
Título: El Discurso del Método
Editorial: AKAL
Año: 2007
N° 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



COMPUTER APPLICATIONS
INFORMATICS

Proper use of common image file formats in handling radiological images

Una efficace dei comuni formati di file immagine nella gestione di immagini radiologiche

S. Facelli · S. Perandini · A. Comi · M. P'Onofrio · R. Paoletti

Stato di Radiologia, Policlinico G.B. Rossi, Università di Verona, P.le L.A. Moro, 13745 Verona, Italy
 Correspondence to: S. Facelli, Tel.: +39-045-832001, Fax: +39-045-832046, e-mail: sfacelli@univr.it
 Received: 11 April 2008 / Accepted: 11 May 2008 / Published online: 17 March 2009
 © Springer-Verlag 2009

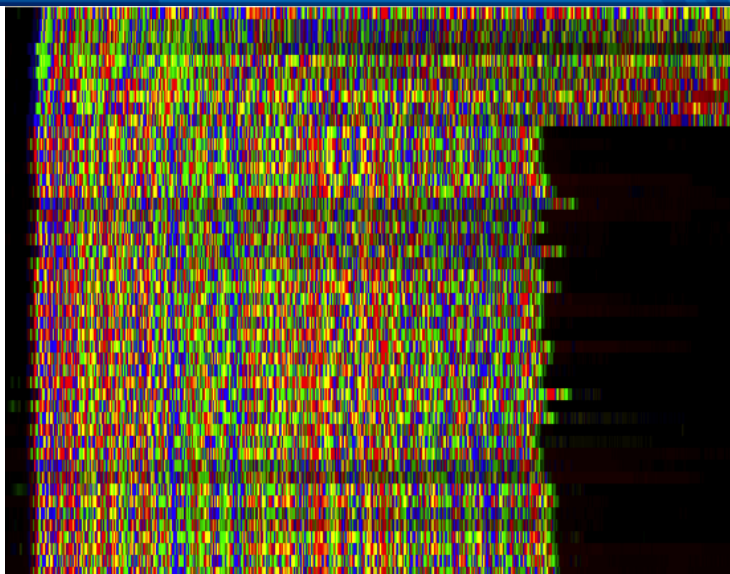
Abstract
 This paper highlights the differences among the most common file formats used for storing digital radiological images. It presents the proper use of these formats in practice and the advantages of handling the data using a practical application in daily radiological practice. The authors provide a simple yet effective introduction to the concept of "file format" and describe the algorithms and data formats of the most common formats (BMP, JPEG, GIF, TIFF, PNG, EPS, PDF) and Portable Network Graphics (PNG).

Keywords Image file format · Radiology image · BMP · JPEG · GIF · TIFF · EPS · PDF · PNG · TIFF

Riassunto
 È illustrata la differenza tra i più comuni formati di salvataggio delle immagini radiologiche digitali. Si descrivono le caratteristiche e le proprietà di questi formati, si presentano le differenze tra di loro e si fornisce un'analisi pratica sul loro utilizzo in radiologia. Gli autori presentano in modo semplice ma esauriente il concetto di "formato file" e gli algoritmi dei formati più comuni (BMP, JPEG, GIF, TIFF, EPS, PDF) e di loro derivazioni principali.

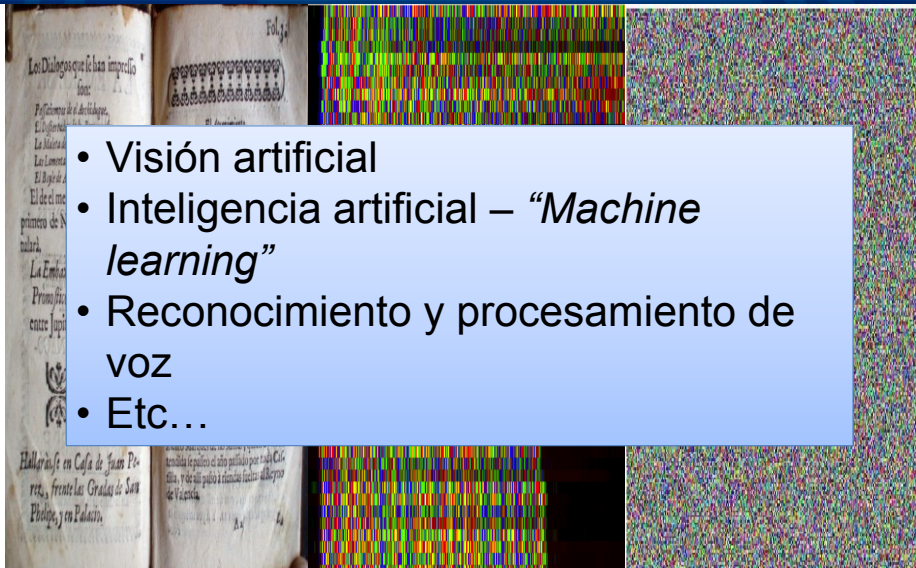
Parole chiave Formato immagine · Immagine radiologica · BMP · JPEG · GIF · TIFF · EPS · PDF · PNG · TIFF

¿Cómo sacar información de los datos desestructurados?





Los datos solo son datos



- Visión artificial
- Inteligencia artificial – “*Machine learning*”
- Reconocimiento y procesamiento de VOZ
- Etc...

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

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



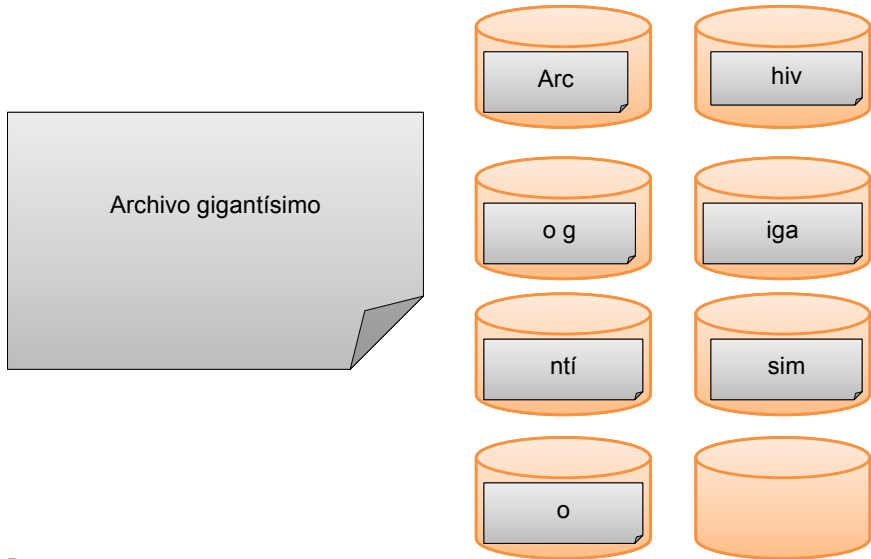
La tecnología detrás del Big Data. Mario Macías. www.xaas.guru

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



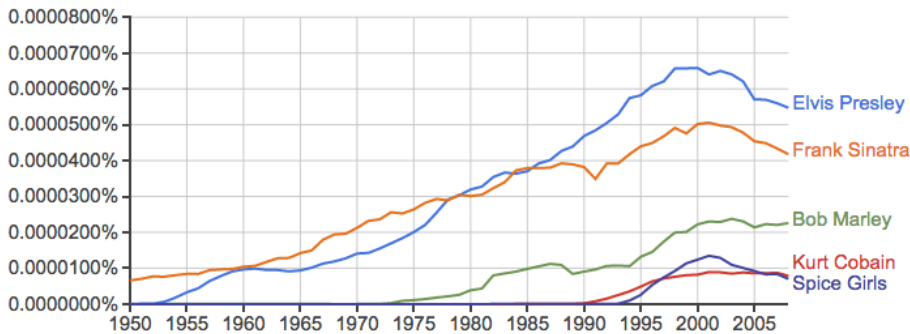
Una pila de software para el Big Data





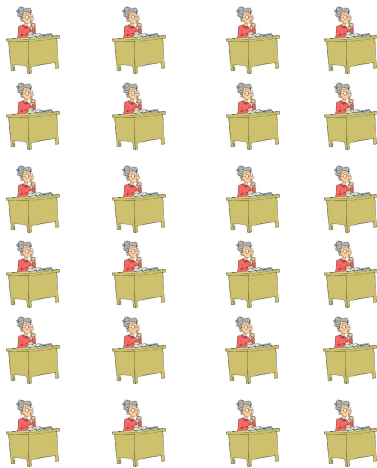
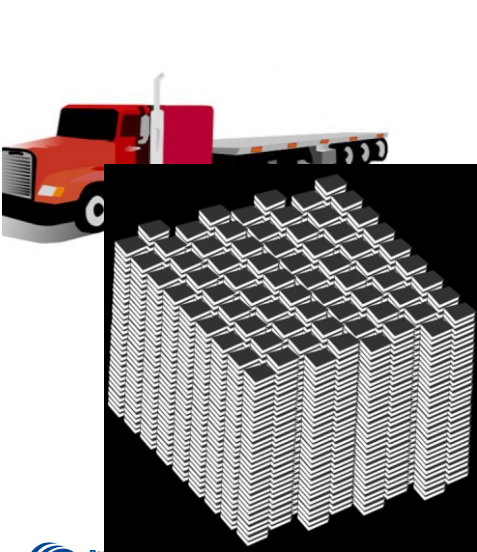
Google books Ngram Viewer

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



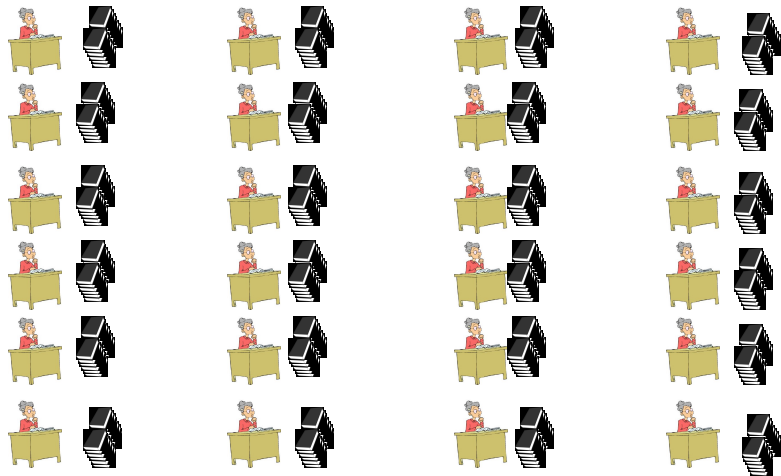
La tecnología detrás del Big Data. Mario Macías. www.xaas.guru

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



La tecnología detrás del Big Data. Mario Macías. www.xaas.guru

Paso 1. Dividir el trabajo



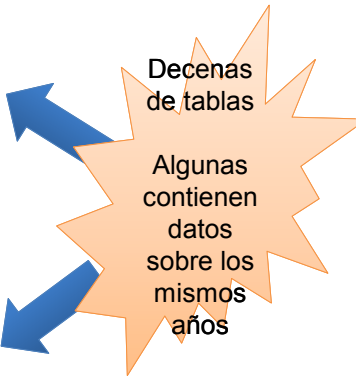
Paso 2. Map



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



Paso 3. Reduce

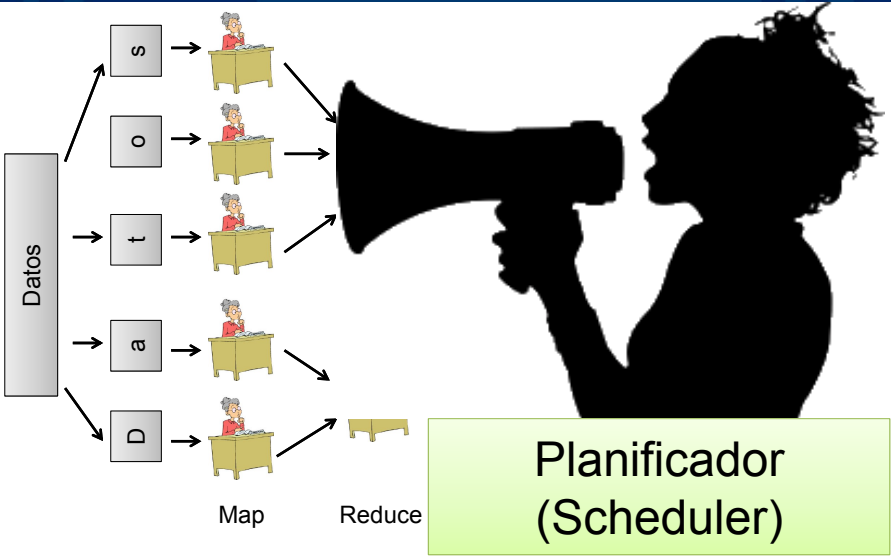
Año Libro	Apariciones
1920	15
1950	22
1955	33
1958	32
1960	40

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

Mario Macías Lloret

Del bit a la Nube



2,99€

amazonkindle

Google play

BARNES
& NOBLE



iBooks

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