

Introducción al Análisis de Datos

Christopher Expósito Izquierdo

Airam Expósito Márquez

Israel López Plata

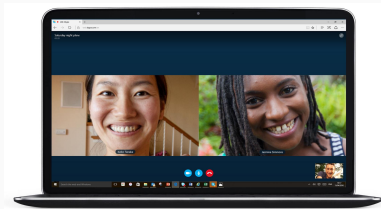
{cexposit, aexposim, ilopezpl}@ull.edu.es



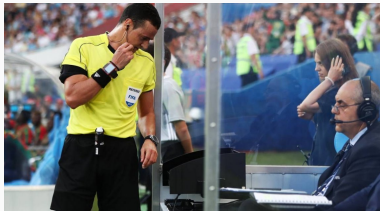
1. Introducción
2. ¿Qué es Big Data?
3. Tratamiento de la Información
4. Tecnologías de Big Data
5. Infraestructura del Big Data

Introducción

Introducción:



Comunicación



Deporte



Entretenimiento



Banca

Introducción:



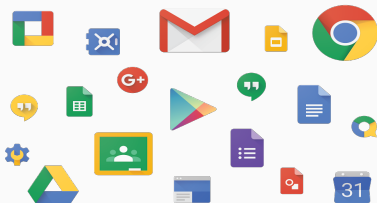
Introducción:



Introducción:



MOOCs



Herramientas de soporte



Herramientas sociales



Comunicación

¿Qué es Big Data?

¿Qué es Big Data?:



¿Qué es Big Data?:



The Internet in Real Time



By the time you finish reading this sentence, there will have been 219,000 new Facebook posts, 22,800 new tweets, 7,000 apps downloaded, and about \$9,000 worth of items sold on Amazon... depending on your reading speed, of course. Now that the Internet is widely available, just one second of global online activity is jam-packed full of events, from communication with others to data storage to entertainment options galore.

For example, in the amount of time you've been on this page, this is how much data has already passed through the Internet.

4,905,600

GIGABYTES OF DATA

The amount of data uploaded to the Internet in a single second is a staggering 24,000 gigabytes. Cisco forecasts that monthly Internet data will reach 81.3 exabytes – or 1 billion gigabytes – by the year 2016, pushing the amount of online activity even higher.

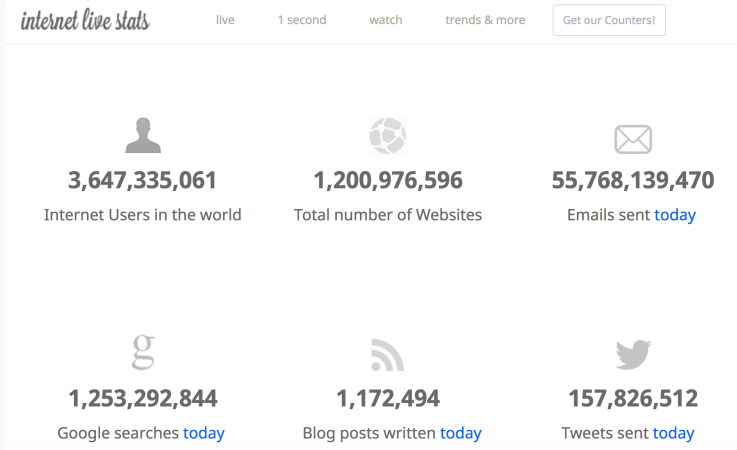
Want to know what else happens in single second online? Scroll down to find out what has happened on the Internet just since you loaded this page!

<https://www.webpagefx.com/internet-real-time/#tweetfx>

- Home
- Facebook
- Twitter
- LinkedIn
- Instagram
- Foursquare
- Pinterest
- Tinder
- WhatsApp
- Snapchat
- Skype
- Amazon
- Google
- Kickstarter
- Yelp
- Email
- Dropbox
- WordPress
- Tumblr
- Flickr
- reddit
- Android
- iOS
- YouTube
- Justify

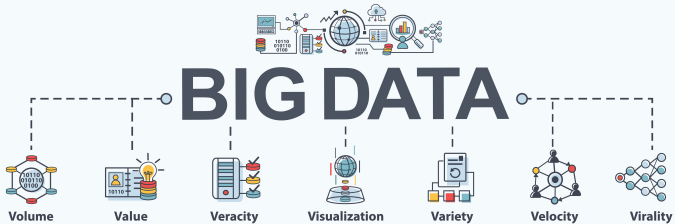
<https://www.webpagefx.com/internet-real-time>

¿Qué es Big Data?:



<http://www.internetlivestats.com>

¿Qué es Big Data?:



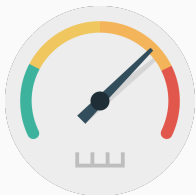
¿Qué es Big Data?:

Volumen:

- Hoy en día se estima que se están produciendo 2.5 quintillones de bytes de datos al día en todo el mundo.
- El 90% de los datos proviene de orígenes que no es posible ser almacenado en bases de datos convencionales.



Velocidad:



- Ratio al que los datos llegan y también cómo de rápido éstos deben ser procesados.
- Existen técnicas para poder inferir conocimiento en tiempo real y responder a consultas sobre flujos de datos.

¿Qué es Big Data?:

Variedad:

- Amplio abanico de tipos de datos que un conjunto de datos particular puede tener.
- Nuevas tecnologías han tenido que ir apareciendo para dar respuesta a la gran variedad de datos existentes.



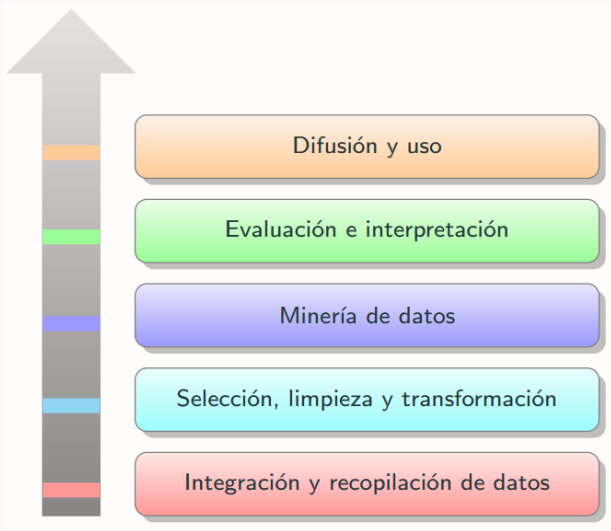
Veracidad:



- Contar con un gran volumen de datos y que sean altamente variados no suele ser suficiente para satisfacer las demandas educativas.
- Si un centro educativo no puede confiar en sus datos, realmente tiene un problema educativo grave.

Tratamiento de la Información

Tratamiento de la Información:



Tratamiento de la Información:



Identificación inmediata de errores dentro del centro

Identificación de estrategias educativas más eficaces



Ahorro de costes

Seguir la evolución de los alumnos



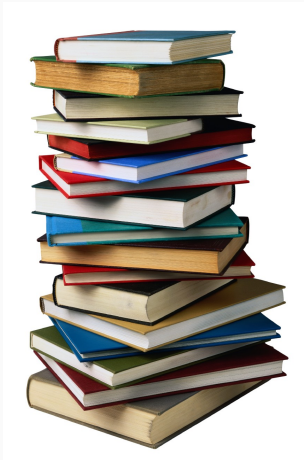
Tecnologías de Big Data

Tecnologías de Big Data:

Los datos son interesantes pero el **valor** real viene cuando se aplican **técnicas de análisis**



Análisis de textos:



- Proceso para la extracción de información y de conocimiento desde grandes volúmenes de textos (informes, correos electrónicos, documento, páginas web, blogs,...)
- El principal inconveniente está en que, aunque son datos estructurados desde si consideramos la legibilidad de un humano, son datos no estructurados considerando la perspectiva de la analítica.

Categorización	De acuerdo a sus atributos, asignar una o varias categorías o clases a cada uno de los documentos disponibles.
Clusterización	Técnicas para crear grupos o clústers de documentos similares.
Extracción de conceptos	Procedimientos para extraer conceptos a partir de textos.
Análisis de sentimientos	Pretende extraer la opinión subjetiva o sentimiento a partir de un texto dado.
Síntesis	Creación automática de un resumen para identificar los aspectos más importantes del texto.

Análisis de audios:



Analizando las interacciones en forma de voz del alumnado a través de herramientas online (por ejemplo, Adobe Connect) se puede identificar qué alumnos están enfadados o frustrados, qué actividades, procesos o áreas mejorar o qué alumnos necesitan formación adicional.

Análisis de imágenes:



- Los últimos avances en análisis de video emplean algoritmos que recorren los vídeos, escena por escena, fotograma a fotograma e identifican realmente qué hay en cada uno de ellos.
- Posteriormente, indexan la información y la usan para identificar patrones o referencias cruzadas con otras herramientas de análisis

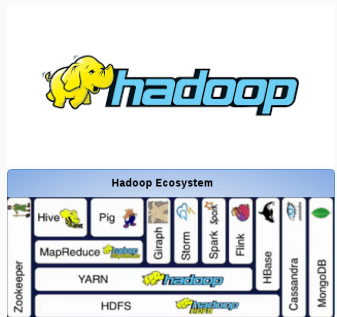
Reconocimiento de caras	Se pretende identificar automáticamente a una persona a partir de una imagen o fotograma en un vídeo.
Análisis de comportamiento	Técnicas para extraer patrones de movimiento.
Análisis de situación	Identificar cambios o anomalías en secuencias de imágenes.

Infraestructura del Big Data

Infraestructura del Big Data:

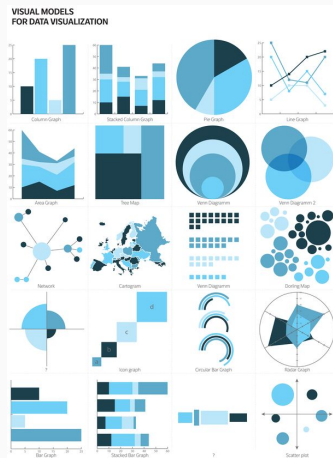
Hadoop es una herramienta basada en código abierto destinada para el almacenamiento, procesamiento y análisis de grandes volúmenes de datos.

- **HDFS**. Es un sistema distribuido de ficheros que hace que un mismo fichero no sea almacenado en un solo equipo.
- **MapReduce**. Es un marco de trabajo destinado al aislamiento del programador en las labores de paralelismo.



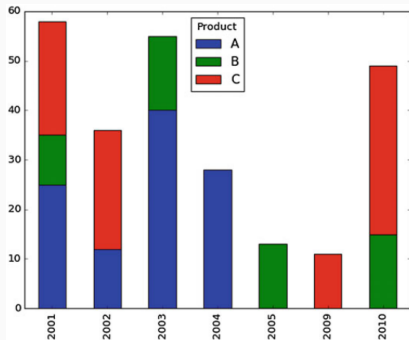
Infraestructura del Big Data:

- La visualización de datos es el proceso de **interpretar** los datos y **presentarlos** en un formato gráfico.
- Se ha vuelto más difícil extraer información crucial de los datos disponibles. Ahí es donde se hace evidente la importancia de la visualización de datos.
- La visualización de datos ayuda a las personas a comprender la importancia de los datos al **resumir** y **presentar** una gran cantidad de datos en un formato simple y **fácil de entender** para poder comunicar la información de manera clara y **efectiva**.

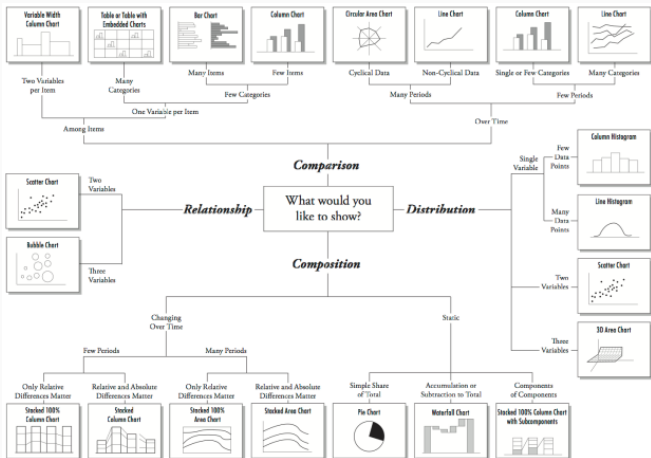


Infraestructura del Big Data:

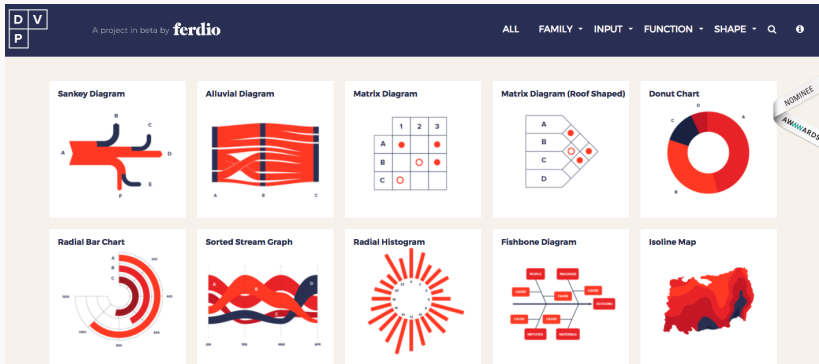
Year	Product	Sales
2001	A	10
2002	A	12
2001	B	4
2003	B	5
2004	A	10
2001	A	11
2001	C	12
2002	C	15
2001	C	11
2002	C	9
2003	A	7
2001	A	4
2001	B	6
2003	B	10
2010	C	11
2010	C	23
2009	C	11
2003	A	12
2004	A	6
2010	B	8
2010	B	7
2003	A	21
2004	A	12
2005	B	13



Infraestructura del Big Data:

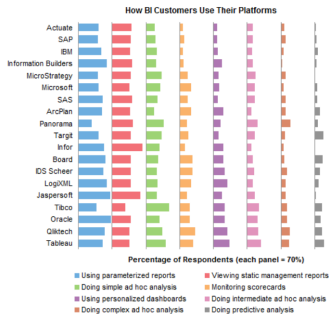
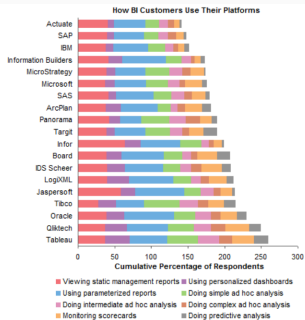


Infraestructura del Big Data:



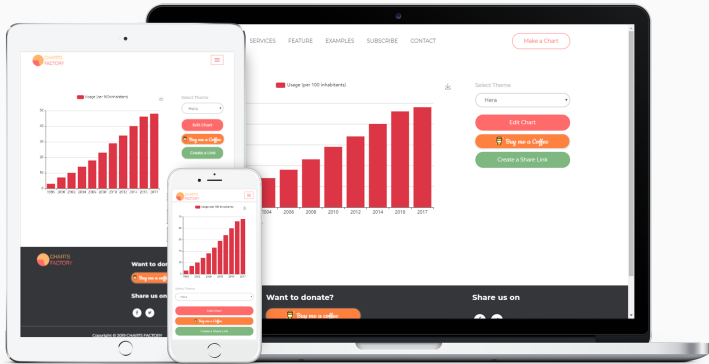
<https://datavizproject.com>

Infraestructura del Big Data:

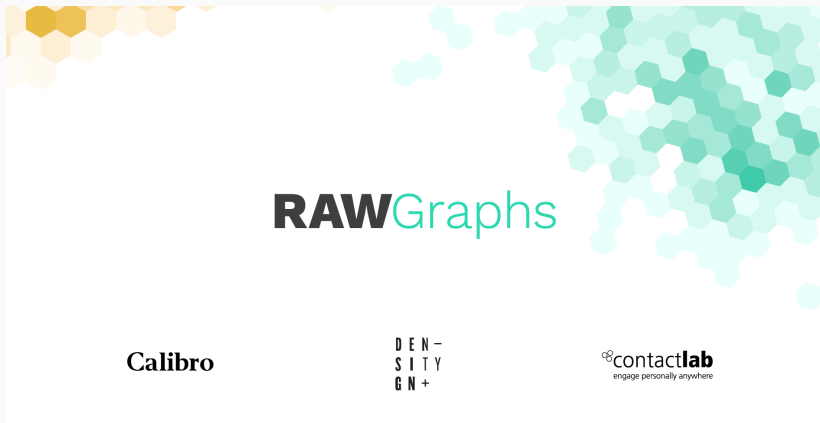


<https://github.com/d3/d3/wiki/Gallery>

Infraestructura del Big Data:

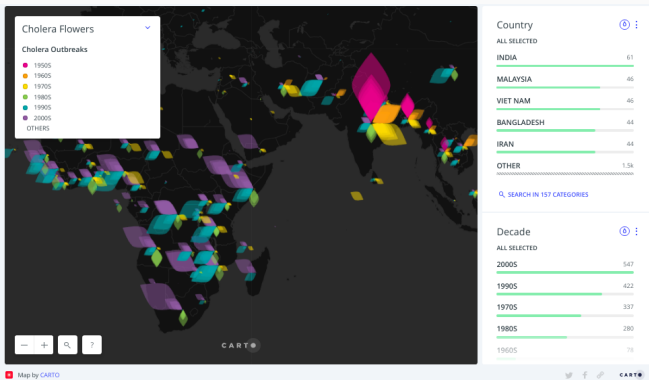


<https://chartsfactory.com>



<https://rawgraphs.io>

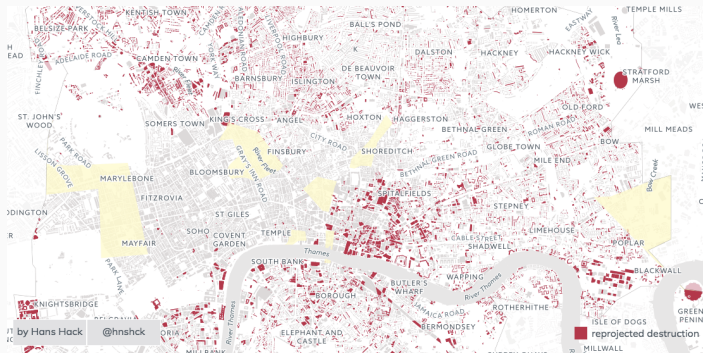
Repositorios:



<https://carto.com/blog/forty-brilliant-open-data-projects-preparing-smart-cities-2018>

Infraestructura del Big Data:

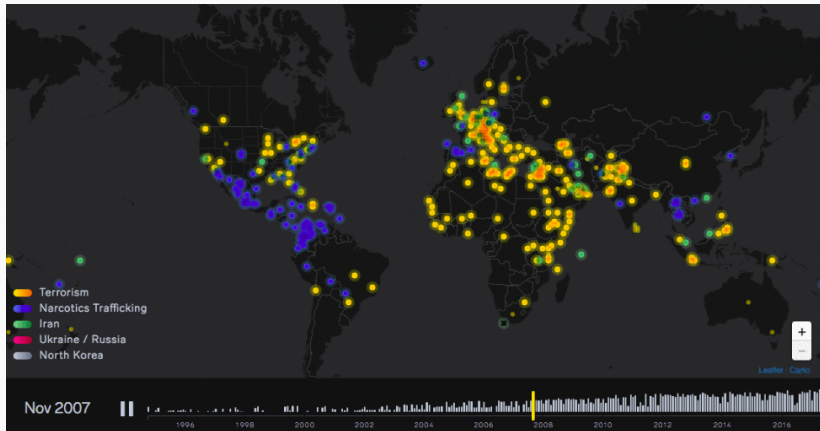
Uso de Mapas:



<https://carto.com/blog/eighty-data-visualizations-examples-using-location-data-maps/>

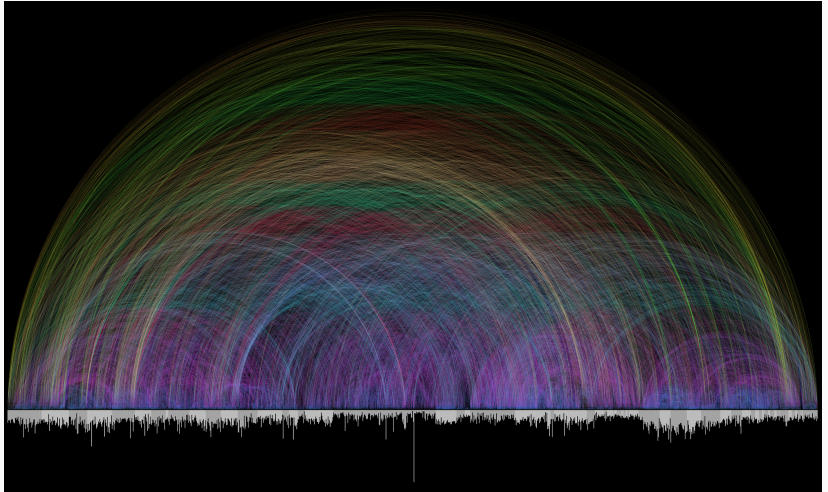
Infraestructura del Big Data:

Stories:

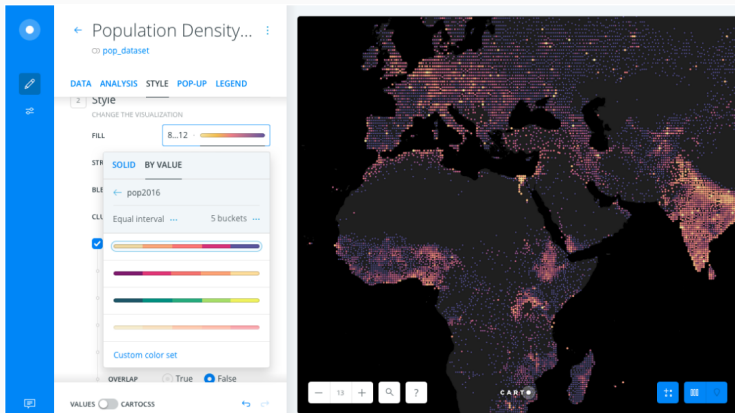


<https://labs.enigma.com/sanctions-tracker/>

Stories:



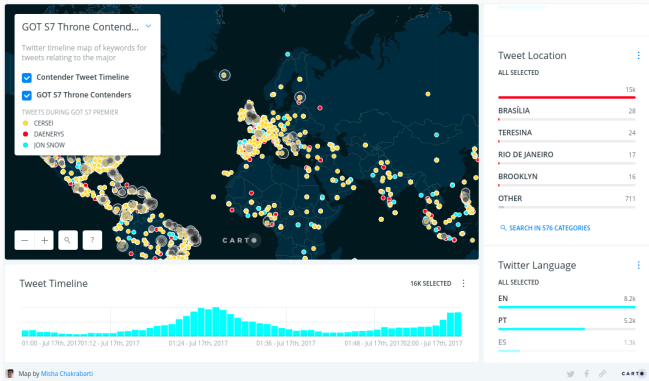
Esquemas de color:



<https://carto.com/carto-colors/>

Infraestructura del Big Data:

Eventos:



<https://youtu.be/KOHdPQU2HQ>