



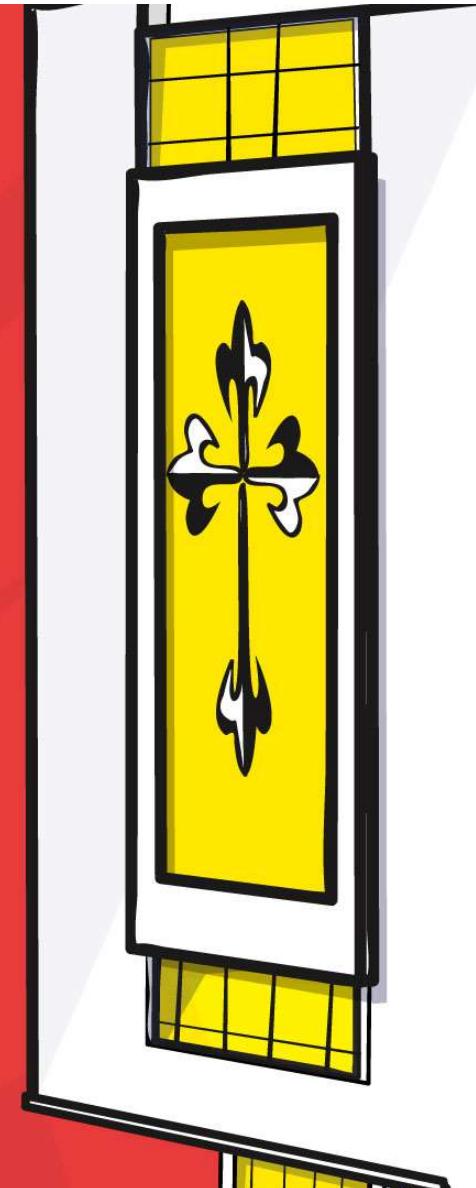
SANTO
Seccional Tunja

www.usntunja.edu.co



BIG DATA

Facultad de Ingeniería de Sistemas



Grupo de Whatsapp



Nube de tags

Ingrese a menti.com y utilice el código 4243 8853 o ingrese por el QR



BIG DATA



“Big data es una colección de grandes volúmenes de datos, complejos y muy difíciles de procesar, a través de herramientas de gestión y procesamiento de datos tradicionales”.

“Tecnología que utilizamos cuando tenemos que manejar un conjunto de información tan grande que no podemos hacer en una sola máquina”.

BIG DATA



DATA SCIENCE

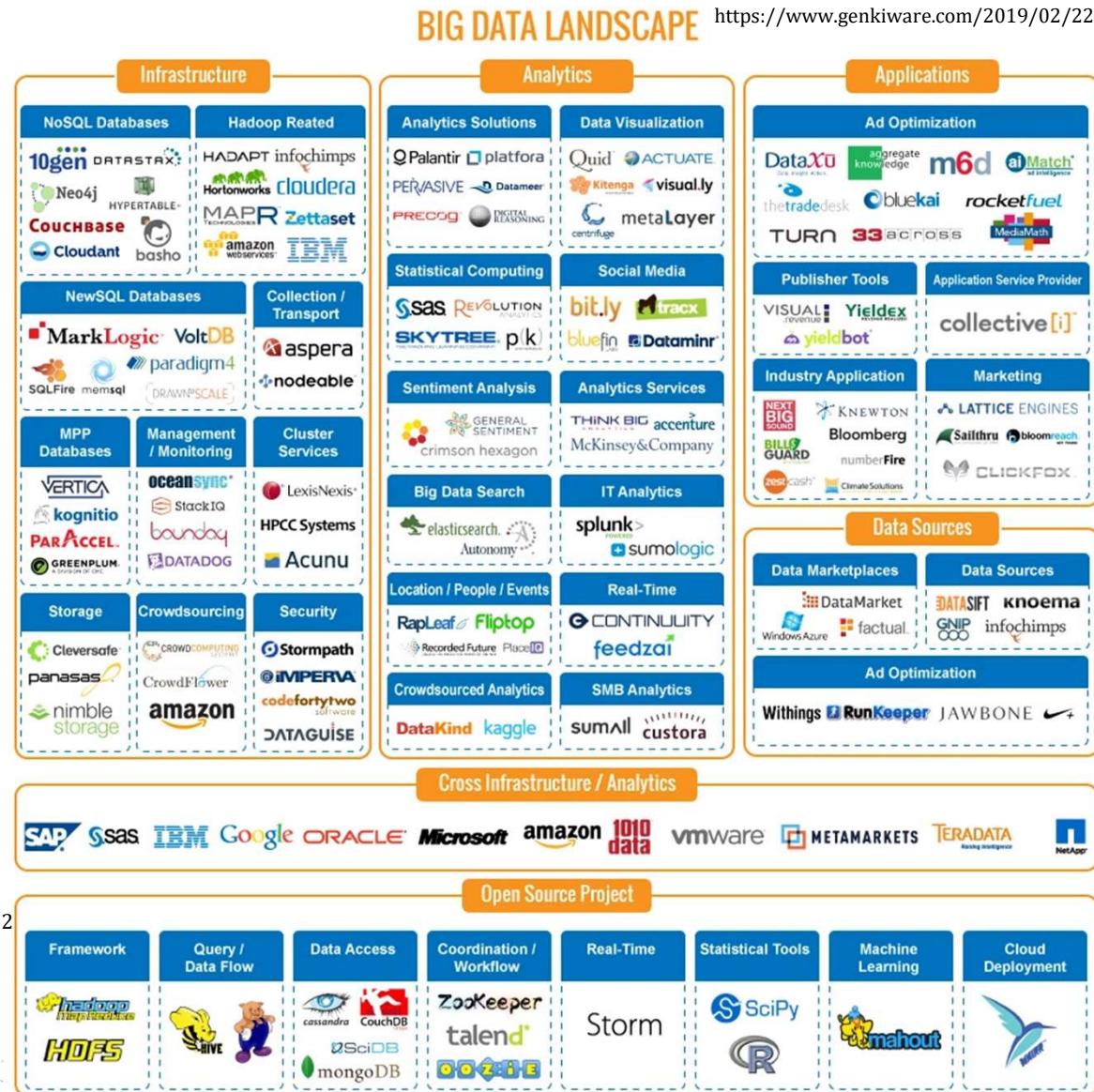
“Campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados”.

Las herramientas de datos con las que trabajamos hoy en día son cada vez mejores.

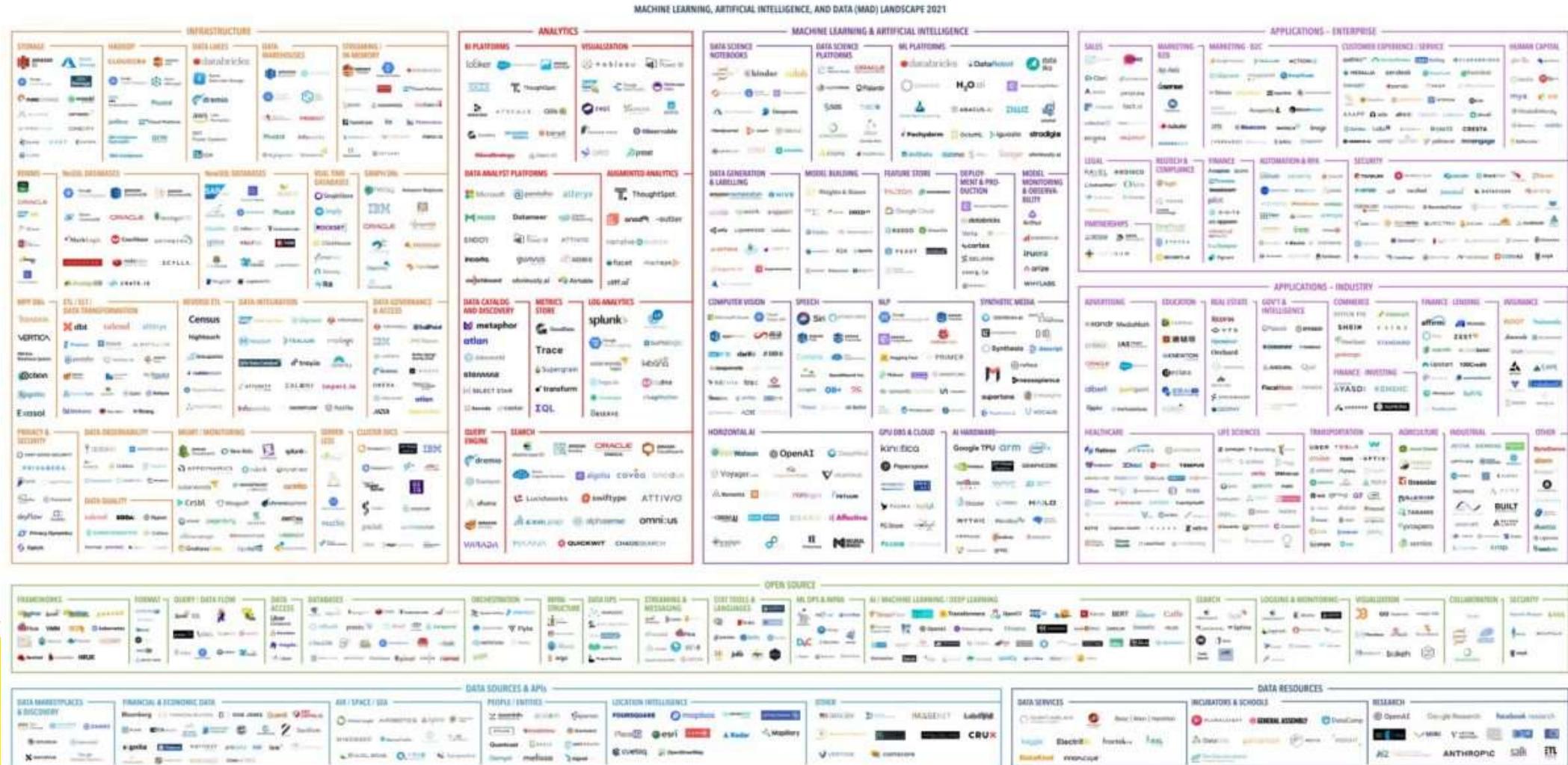
En un entorno en el que las tareas de datos específicas que solían ser realmente difíciles se simplifican o incluso se automatizan por completo con buenas herramientas, puede ser difícil dar un paso atrás y apreciar que la "pila de datos moderna" en realidad está creciendo en tamaño y complejidad.

Eric Dodds

Jefe de marketing de productos en RudderStack



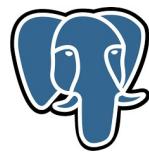
<https://www.rudderstack.com/blog/the-state-of-data-engineering-in-2022>



Herramientas en Big Data

Bases de datos relacionales

ORACLE



Tradicionales



Almacenamiento

Sistema de archivos distribuidos

No SQL

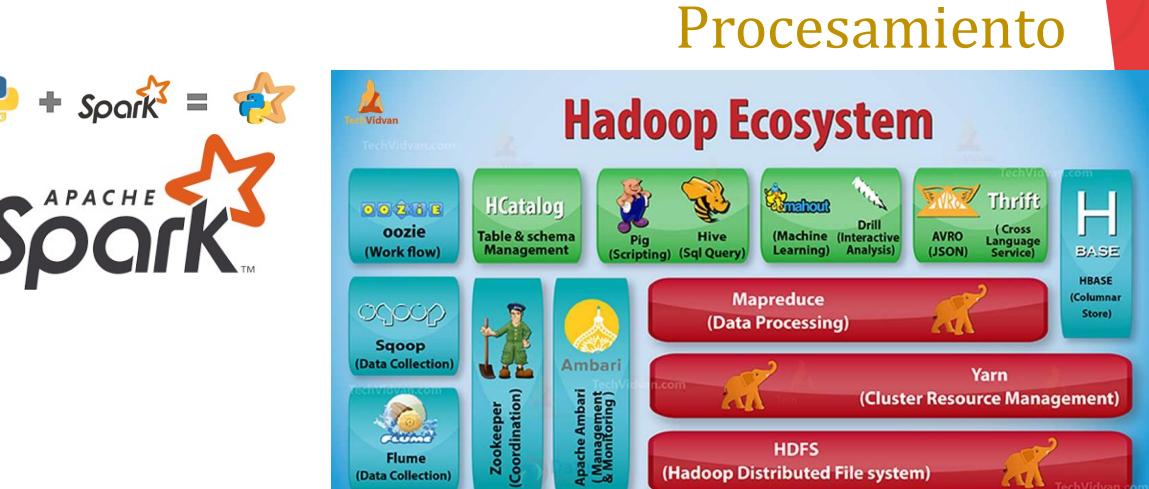
 **mongoDB**®

 **hadoop**
HDFS

 **cassandra**

Big Data

 **APACHE Spark™**



Big Data y Ciencia de los Datos

Big Data: Se refiere a enormes volúmenes de datos que no pueden procesarse de manera efectiva con las aplicaciones tradicionales que actualmente se aplican. De acuerdo con la guía de *Amazon Web Service*, se considera al *Big Data* como una colección considerable de datos con dificultades para almacenarse en bases de datos tradicionales, y también para procesarse en servidores estándar y para analizarse con aplicaciones habituales.



El término se suele relacionar con ciencia de datos, pues esa suele ser su fuente de información para análisis; la ciencia de datos logra analizar los grandes conjuntos de datos desordenados e incompletos, para llegar a hallazgos que impulsan decisiones sobre operaciones y productos.

Big Data e Inteligencia de negocios

La inteligencia de negocios (BI) combina análisis de negocios, minería de datos, visualización de datos, herramientas e infraestructura de datos, y las prácticas recomendadas para ayudar a las organizaciones a tomar decisiones más basadas en los datos.

En la práctica, se tiene una inteligencia de negocios cuando hay una visión integral de los datos de la organización y se utilizan para impulsar el cambio, eliminar las inefficiencias y adaptarse rápidamente a los cambios del mercado o del suministro.



<https://todobi.com/herramientas-business-intelligence-open-source/>

Habilidades duras y blandas

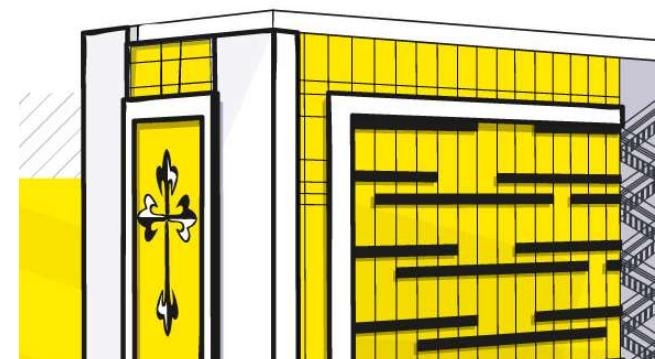
HABILIDADES DURAS



HABILIDADES BLANDAS



Desarrollado por la Dra Nuria Vanegas



HABILIDADES BLANDAS

Habilidades relacionadas a cómo trabajas e interactúas con otros en el trabajo



Liderazgo y administración



Pensamiento crítico



Pensamiento de diseño
y Creatividad



Solución a problemas



Innovación y disruptión



HABILIDADES DURAS

Habilidades específicas del trabajo y conocimientos técnicos medibles



Inteligencia artificial



Análisis y Ciencia de Datos



Aprendizaje automático



Marketing digital



Investigación de mercados e
información al consumidor

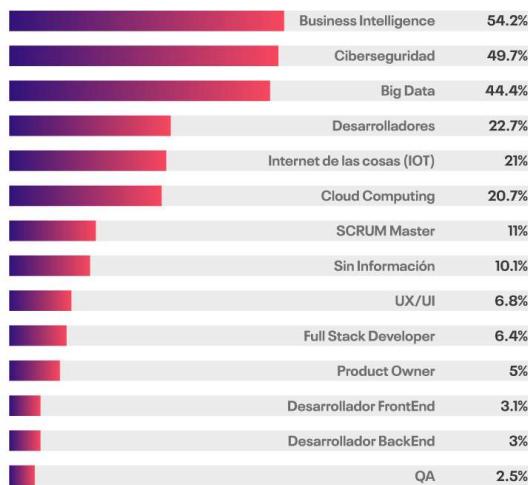


Oportunidades laborales para el 2022

Según el último informe de la consultora Michael Page, los sectores que seguirán teniendo el mayor crecimiento este año son el sector tecnología y salud. El estudio se basó en encuestas realizadas a la alta dirección de diversas empresas de Latinoamérica.

En este sondeo tuvieron la opción de elegir los 3 perfiles con mayor demanda dentro de sus empresas. El 54.2% señaló el Business Intelligence como el perfil más requerido, seguido por Ciberseguridad con un 49.7% y Big data con 44.4%.

¿CUÁLES CREEES QUE SERÁN LOS **PERFILES DE TECNOLOGÍA MÁS DEMANDADOS EN 2022?**

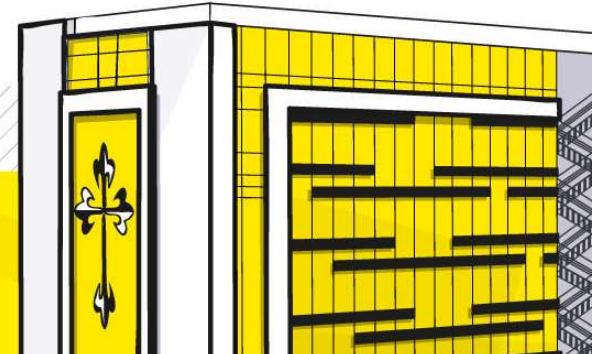


Fuente: Michael Page



<https://blog.continua.edu.pe/habilidades-blandas>

<https://www.michaelpage.com.co/>



Cuánta información se genera cada día

Byte es un término creado por Werner Buchholz en 1957 como una **unidad de información digital** equivalente a cuatro bits (*binary digit*, dígito binario) originalmente y posteriormente como estándar se adoptó que 1 byte equivale a ocho bits. La palabra byte proviene de *bite*, que significa mordisco, como la cantidad más pequeña de datos que un ordenador podía "morder" a la vez.

- 1 byte = 8 bits
- 1 kilobyte (kB o kbyte) = 1024 bytes
- 1 megabyte (MB o Mbytes) = 1024 kilobytes
- 1 gigabyte (GB o Gbytes) = 1024 megabytes
- 1 terabyte (TB o Tbytes) = 1024 gigabytes
- 1 petabyte (PB o Pbytes) = 1024 terabytes
- 1 exabyte (EB o Ebytes) = 1024 petabytes
- 1 zettabyte (ZB o Zbytes) = 1024 exabytes
- 1 yottabyte (YB o Ybytes) = 1024 zettabytes

Prefijo	Símbolo del prefijo	Nombre resultante del prefijo + byte	Símbolo del múltiplo del byte	Factor y valor en el SI
Valor de referencia		byte	B	$10^0 = 1$
kilo	k	kilobyte	kB	$10^3 = 1\ 000$
mega	M	megabyte	MB	$10^6 = 1\ 000\ 000$
giga	G	gigabyte	GB	$10^9 = 1\ 000\ 000\ 000$
tera	T	terabyte	TB	$10^{12} = 1\ 000\ 000\ 000\ 000$
peta	P	petabyte	PB	$10^{15} = 1\ 000\ 000\ 000\ 000\ 000$
exa	E	exabyte	EB	$10^{18} = 1\ 000\ 000\ 000\ 000\ 000\ 000$
zetta	Z	zettabyte	ZB	$10^{21} = 1\ 000\ 000\ 000\ 000\ 000\ 000\ 000$
yotta	Y	yottabyte	YB	$10^{24} = 1\ 000\ 000\ 000\ 000\ 000\ 000\ 000\ 000$

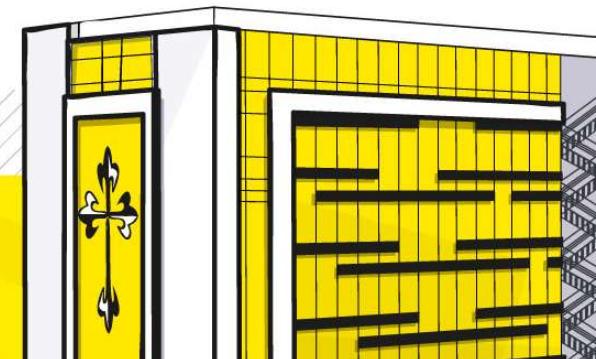
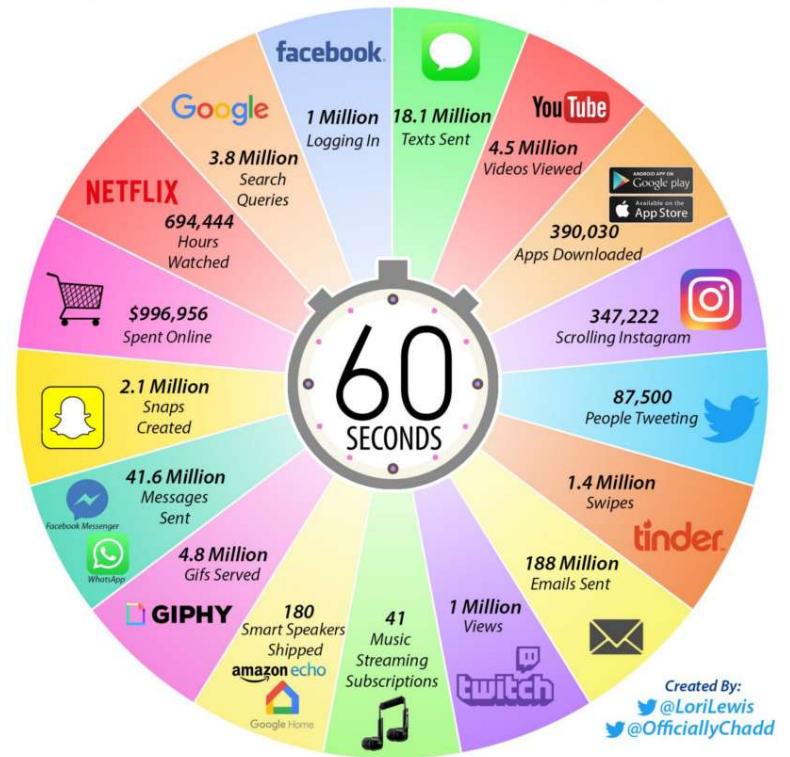
Prefijo	Símbolo del prefijo	Nombre resultante del prefijo + byte	Símbolo del múltiplo del byte	Factor y valor en el ISO/IEC 80000-13
Valor de referencia		byte	B	$2^0 = 1$
kibi	Ki	kibibyte	KiB	$2^{10} = 1\ 024$
mebi	Mi	mebibyte	MiB	$2^{20} = 1\ 048\ 576$
gibi	Gi	gibibyte	GiB	$2^{30} = 1\ 073\ 741\ 824$
tebi	Ti	tebibyte	TiB	$2^{40} = 1\ 099\ 511\ 627\ 776$
pebi	Pi	pebibyte	PiB	$2^{50} = 1\ 125\ 899\ 906\ 842\ 624$
exbi	Ei	exbibyte	EiB	$2^{60} = 1\ 152\ 921\ 504\ 606\ 846\ 976$
zebi	Zi	zebibyte	ZiB	$2^{70} = 1\ 180\ 591\ 620\ 717\ 411\ 303\ 424$
yobi	Yi	yobibyte	YiB	$2^{80} = 1\ 208\ 925\ 819\ 614\ 629\ 174\ 706\ 176$



Cuánta información se genera cada día

Algunas estadísticas diarias clave destacadas en la infografía:

- ✓ Se envían 500 millones de tuits
- ✓ Se envían 294 mil millones de correos electrónicos
- ✓ Se crean 4 petabytes de datos en Facebook
- ✓ Se crean 4 terabytes de datos de cada coche conectado
- ✓ Se envían 65 mil millones de mensajes en WhatsApp
- ✓ Se realizan 5 mil millones de búsquedas
- ✓ Para 2025, se estima que se crearán 463 exabytes de datos cada día en todo el mundo, ¡eso es el equivalente a 212 765 957 DVD por día!



Ejercicio No. 1: Conversión de unidades de almacenamiento

<http://extraconversion.com/es/almacenamiento-de-datos>

1. Complete:

- a) 2 KB representan..... bytes o bits.
- b) 96 bits equivalen a caracteres.
- c) 8 MB es igual a Kbytes.
- d) 3 GB es igual a Mbytes.

2. Los siguientes valores indican distintos tamaños o pesos de información almacenada.
¿cuál es el menor y cuál es el mayor?

- a) 1.576.648 bytes
- b) 1,2 MB
- c) 1.675 KB

3. Un reproductor de MP3 tiene 1 GB de capacidad y se desea almacenar en él archivos de música que tienen un tamaño promedio de 3 MB. ¿Cuántas canciones se pueden guardar?

4. ¿Cuántas fotos podría almacenar una cámara digital con memoria interna de 2 GB si cada foto tiene un tamaño de 2MB?

5. Un pendrive con una capacidad de 1 GB tiene el 25% del espacio libre, ¿podrá almacenar un mapa digitalizado de 280.000 KB? Realice los cálculos.

6. Se calcula que Gmail tiene unos 50 millones de usuarios y se supone que cada uno requiere un almacenamiento de 2747 MB. Estime el tamaño necesario para mantener este servicio. Exprese el resultado en Peta bytes.

7. Un estudio reciente reveló que durante el 2006 la cantidad de información digital creada, capturada y replicada en todo el mundo fue de 161.000 millones de gigabyte, 3 millones de veces la información contenida en todos los libros escritos. ¿A cuántos exabytes corresponde?

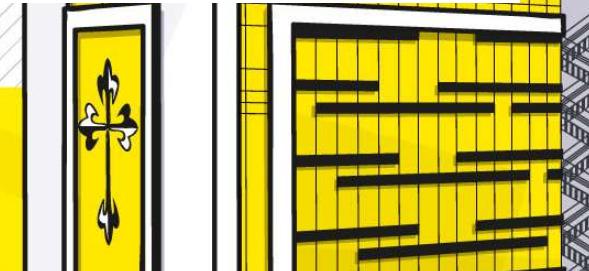
8. La sección de lectores de un diario de la ciudad impone como única restricción para la publicación de las cartas, que el texto no supere los 1500 caracteres. ¿Cuál será el tamaño en KB de un archivo txt que contenga ese texto?

9. Un disco posee una capacidad de almacenamiento de 3.276,80 MB y su espacio utilizado 1 GB. ¿Cuál es la cantidad de bytes libres en dicho disco?

10. Se dispone de un e-book reader con capacidad para almacenar hasta 2GB. Se desea cargar en el mismo los apuntes de las 3 materias del primer cuatrimestre: los de Matemática I pesan 26MB, los de Introducción a la Informática 8.363 KB y los de Sociología de las Organizaciones 3.638.336 bytes. ¿Es posible? Si es así, ¿cuántos KB libres quedarán?

11. Un proveedor de Internet brinda un espacio web de 25 MB. Se elaboró un sitio web que incluye 8 páginas HTML que suman 96.201 bytes en total, 100KB de imágenes y algunas fotos que ocupan 15,4 MB. ¿Cuánto espacio quedará disponible luego de subir el sitio?

12. En la unidad de disco duro (Disco local (D:)) con una capacidad de 49 Gigabyte le insertó una carpeta de videos con un volumen de información de 17000 Megabyte, ¿Cuánto espacio del Disco local (D:) queda disponible?, la respuesta en Kilobyte.



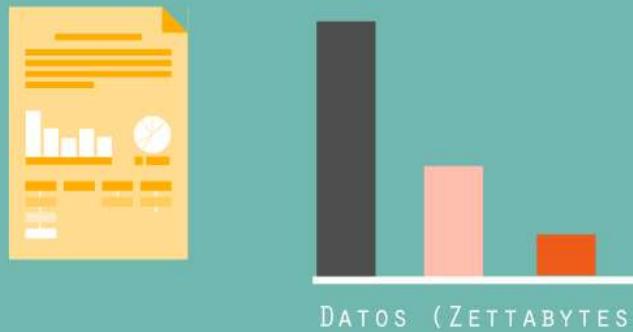
Evolución del Big Data

Evolución de Big Data



Evolución del Big Data

POR DÍA SE CREAN 5.2 ZETTABYTES DE DATOS
Y SE ESPERA QUE ESTE NÚMERO CREZCA A
44 ZETTABYTES PARA EL AÑO 2020.



EN EL 2014, EL UNIVERSO
DIGITAL EQUIVALÍA A 1.7
MEGABYTES POR MINUTO POR
CADA HABITANTE DE LA TIERRA.



EL 17% DE LOS DATOS
SE CREA DESDE PLATAFORMAS MÓVILES.

1

2

La Evolución de Big Data:

CREACIÓN DE DATOS: UNA PERSONA CREA UN DATO
QUE SE ALMACENA EN VARIOS MEDIOS.



6

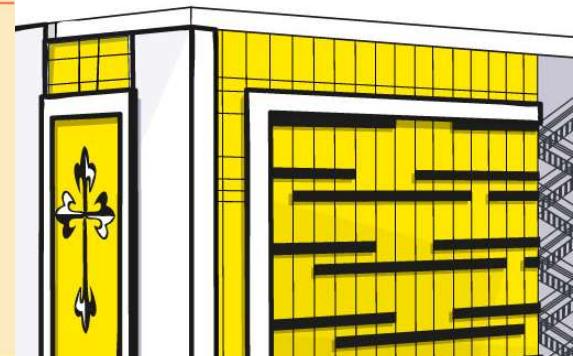
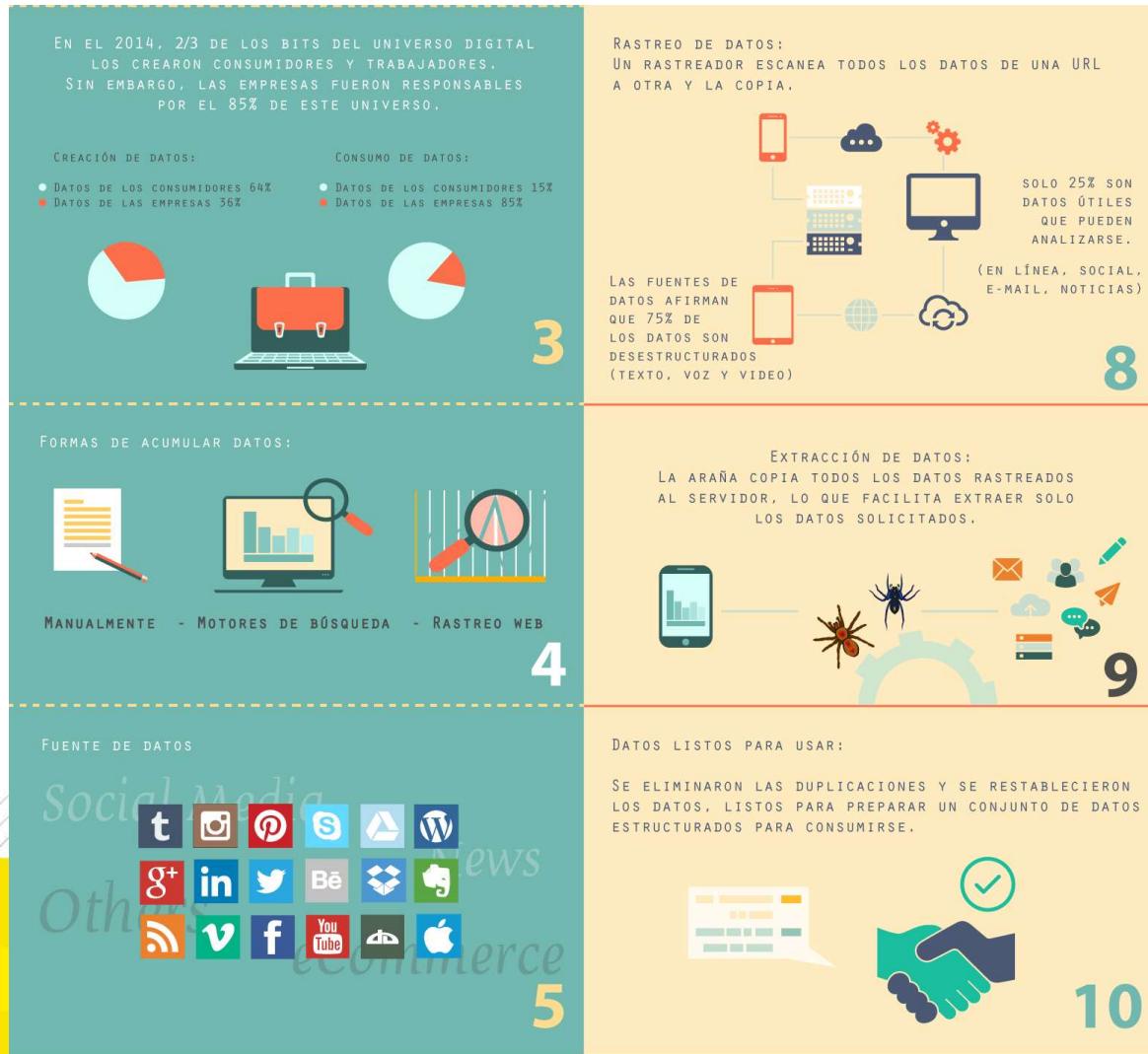
RECOPILACIÓN DE DATOS:
SE RASTREAN, INDEXAN Y RECOPILAN TODOS LOS DATOS
A TRAVÉS DE ARAÑAS WEB (WEB SPIDERS).



7

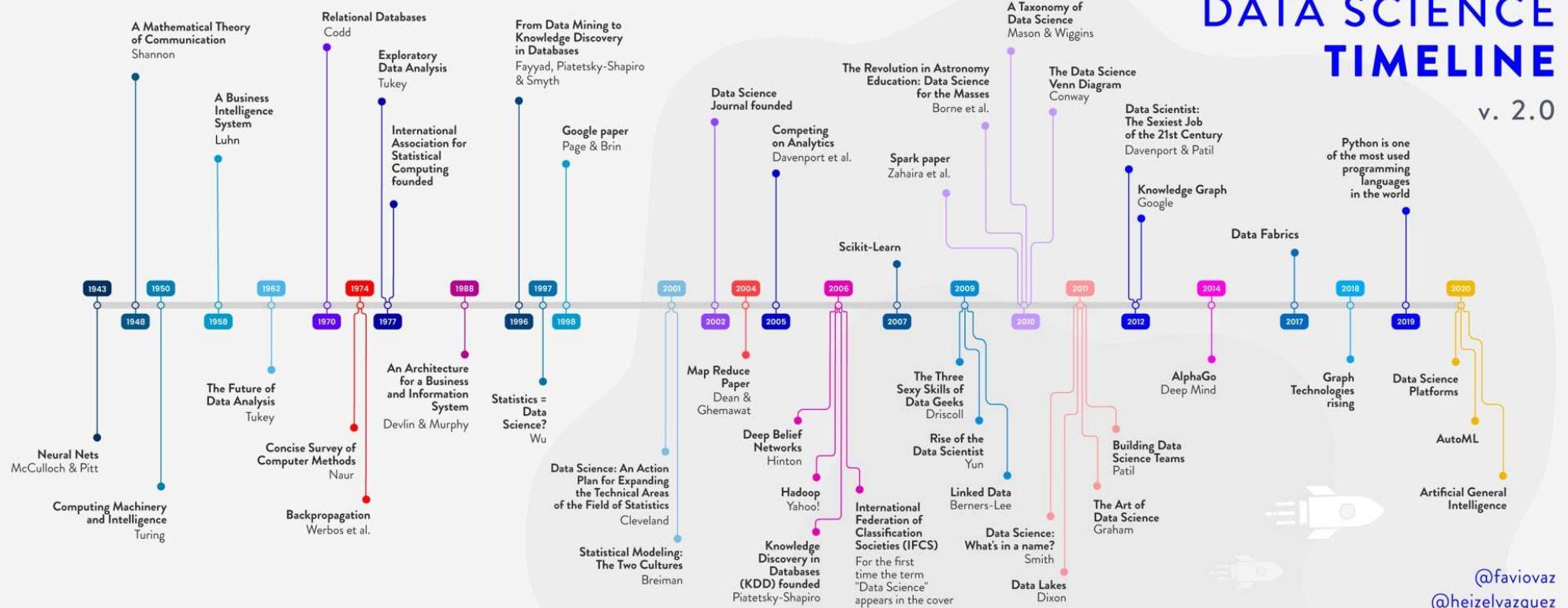


Evolución del Big Data



Evolución de la Ciencia de los Datos

DATA SCIENCE TIMELINE v. 2.0



<https://twitter.com/FavioVaz/status/1258857493607985152/photo/1>

Tipos de datos

Datos Estructurados:

La gran mayoría de las fuentes de datos tradicionales son originadas por datos del tipo estructurados, datos con formato o esquema fijo, que poseen campos fijos y bien definidos.

	nombre	color	edad	altura	peso	puntuacion
1:	Paco	Rojo	24	182	74.8	83
2:	Juan	Green	30	170	70.1	500
3:	Andres	Amarillo	41	169	60.0	20
4:	Natalia	Green	22	183	75.0	865
5:	Vanesa	Verde	31	178	83.9	221
6:	Miriam	Rojo	35	172	76.2	413
7:	Juan	Amarillo	22	164	68.0	902

Ejemplo de datos estructurados



<https://twitter.com/FavioVaz/status/1258857493607985152/photo/1>

Tipos de datos

Datos no Estructurados:

Son las estructuras de datos más difíciles de manejar, podemos encontrar entre los datos no estructurados más conocidos:

- Documentos PDF o Word.
- Audios y videos.
- Correos electrónicos.
- Ficheros multimedia de imagen.
- Artículos y textos, entre otros.

Que trata de la condición y ejercicio del famoso hidalgo D. Quijote de la Mancha

En un lugar de la Mancha, de cuyo nombre no quiero acordarme, no ha mucho tiempo que vivía un hidalgo de los de lanza en astillero, adarga antigua, rocín flaco y galgo corredor. Una olla de algo más vaca que carnero, salpicón las más noches, duelos y quebrantos los sábados, lentejas los viernes, algún palomino de añadura los domingos, consumian las tres partes de su hacienda. El resto della concluían sayo de velarte, calzas de velludo para las fiestas con sus pantuflos de lo mismo, los días de entre semana se honraba con su vellori de lo más fino. Tenía en su casa una ama que pasaba de los cuarenta, y una sobrina que no llegaba a los veinte, y un mozo de campo y plaza, que así ensillaba el rocín como tomaba la podadera. Frisaba la edad de nuestro hidalgo con los cincuenta años, era de compleción recia, seco de carnes, enjuto de rostro; gran madrugador y amigo de la caza. Quieren decir que tenía el sobrenombre de Quijada o Quesada (que en esto hay alguna diferencia en los autores que deste caso escriben), aunque por conjeturas verosímiles se deja entender que se llama Quijana; pero esto importa poco a nuestro cuento; basta que en la narración dél no se salga un punto de la verdad.

Ejemplo de datos no estructurados



<https://twitter.com/FavioVaz/status/1258857493607985152/photo/1>

Tipos de datos

Datos Semi - Estructurados:

Son un híbrido entre los datos estructurados y los datos no estructurados, podríamos decir entonces de manera sencilla, que no presentan una estructura perfectamente definida como los datos estructurados, pero sí presentan una organización definida en sus metadatos donde describen los objetos y sus relaciones.

```
{  
  "marcadores": [  
    {  
      "latitude": 40.416875,  
      "longitude": -3.703308,  
      "city": "Madrid",  
      "description": "Puerta del Sol"  
    },  
    {  
      "latitude": 40.417438,  
      "longitude": -3.693363,  
      "city": "Madrid",  
      "description": "Paseo del Prado"  
    },  
    {  
      "latitude": 40.407015,  
      "longitude": -3.691163,  
      "city": "Madrid",  
      "description": "Estación de  
Atocha"  
    }  
  ]  
}
```

Ejemplo de datos semiestructurados



<https://twitter.com/FavioVaz/status/1258857493607985152/photo/1>



www.usatunja.edu.co

♥ Q Ⓜ @SANTOTOLASTUNJA