

Big Data

Bienvenido - Welcome - Witam



Juanmi Taboada

<http://www.juanmitaboada.com>



@juanmitaboada

<https://es.linkedin.com/in/juanmitaboada>



Juan Miguel Taboada Godoy (1980 - ...)

1996 – Primer ordenador y primera LAN (cable coaxial)

1999 – Universidad de Málaga y **Linux Málaga**

2001 – Grupo de investigación **GEB.uma.es** (4 años)
Cluster computación +20 nodos (**OpenMosix**)

2002 – Presidente de Asociación **Málaga Wireless**

2003 – Beca en **Neurociencia** en SUNY

Teleruta (Ministerio de Fomento – 2 años)

2005 – Autónomo:

- Nace **Centrologic**

- Polonia (2 años)

- Likindoy (Axaragua + **Junta Andalucía**)

2008 – Responsable Sistemas en PontGrup

2011 – Adquisición Datos en Bética Fotovoltáicas

2012 – **SAFECLON** y SCRUM/KANBAN

2013 – **Executive MBA**

2014 – **Aeronáutica: Django + AngularJS**

2015 – **Industria: Likindoy + Big Data**





¿Información?

Información

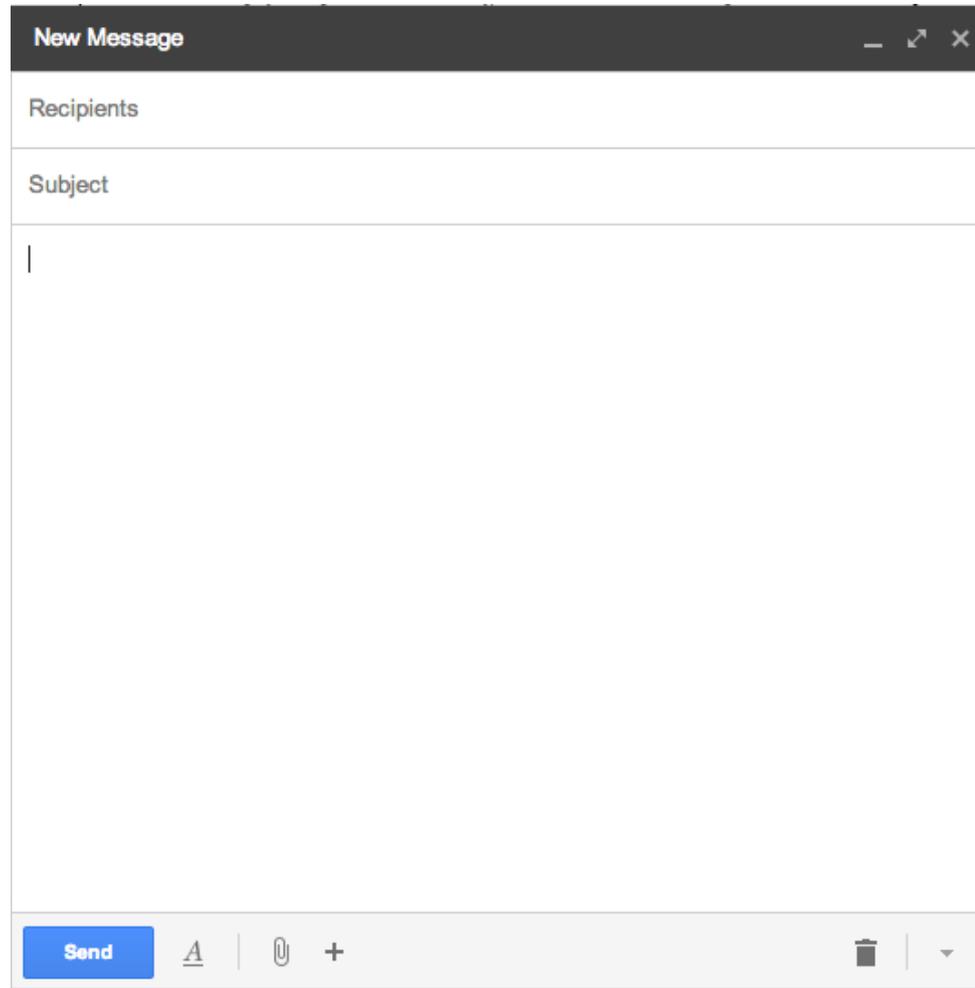
es un conjunto organizado de datos procesados,
que constituyen un mensaje
que cambia el estado de conocimiento
del sujeto o sistema que recibe dicho mensaje.

(fuente Wikipedia)

Información digital



Información digital → Email



A screenshot of a 'New Message' email composition window. The window has a dark title bar with the text 'New Message' and standard window controls (minimize, maximize, close). Below the title bar, there are three main sections: 'Recipients', 'Subject', and a large text area for the message body. The 'Recipients' and 'Subject' fields are currently empty. The text area contains a single vertical cursor. At the bottom of the window, there is a toolbar with a blue 'Send' button, a text formatting icon (underline), an attachment icon, a plus sign, a trash icon, and a dropdown arrow.

Información digital → Email



New Message

Recipients

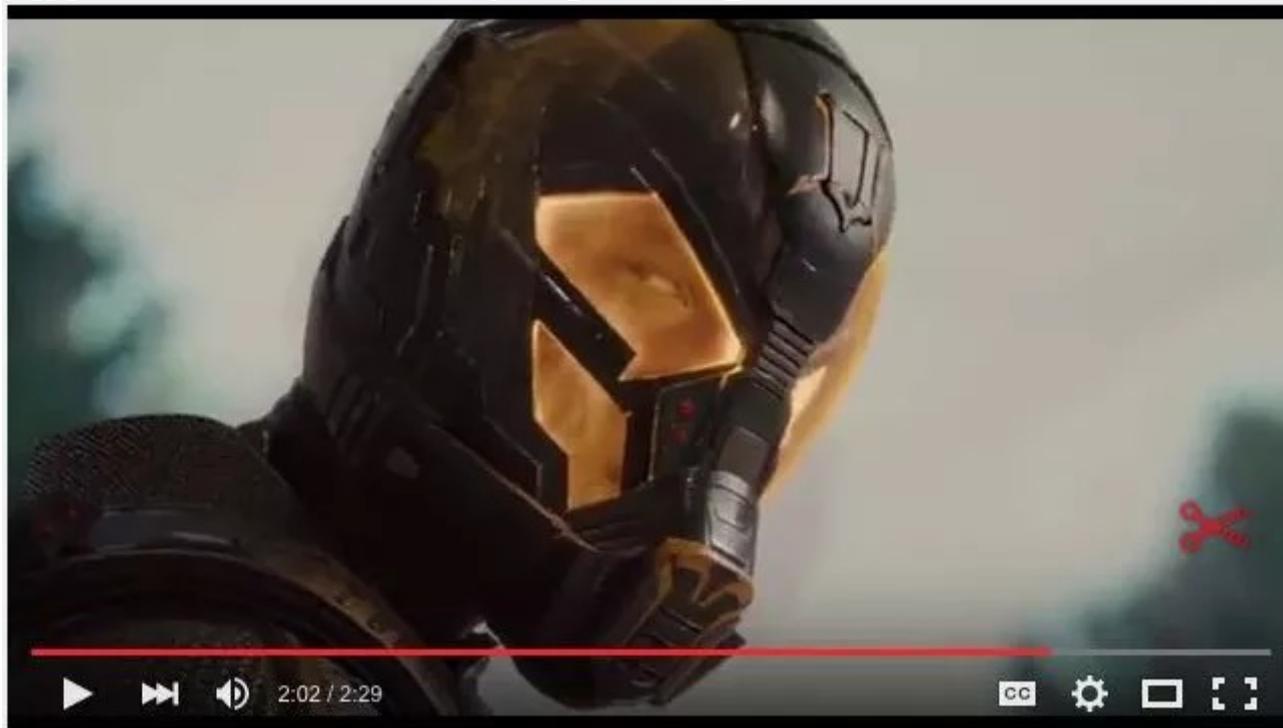
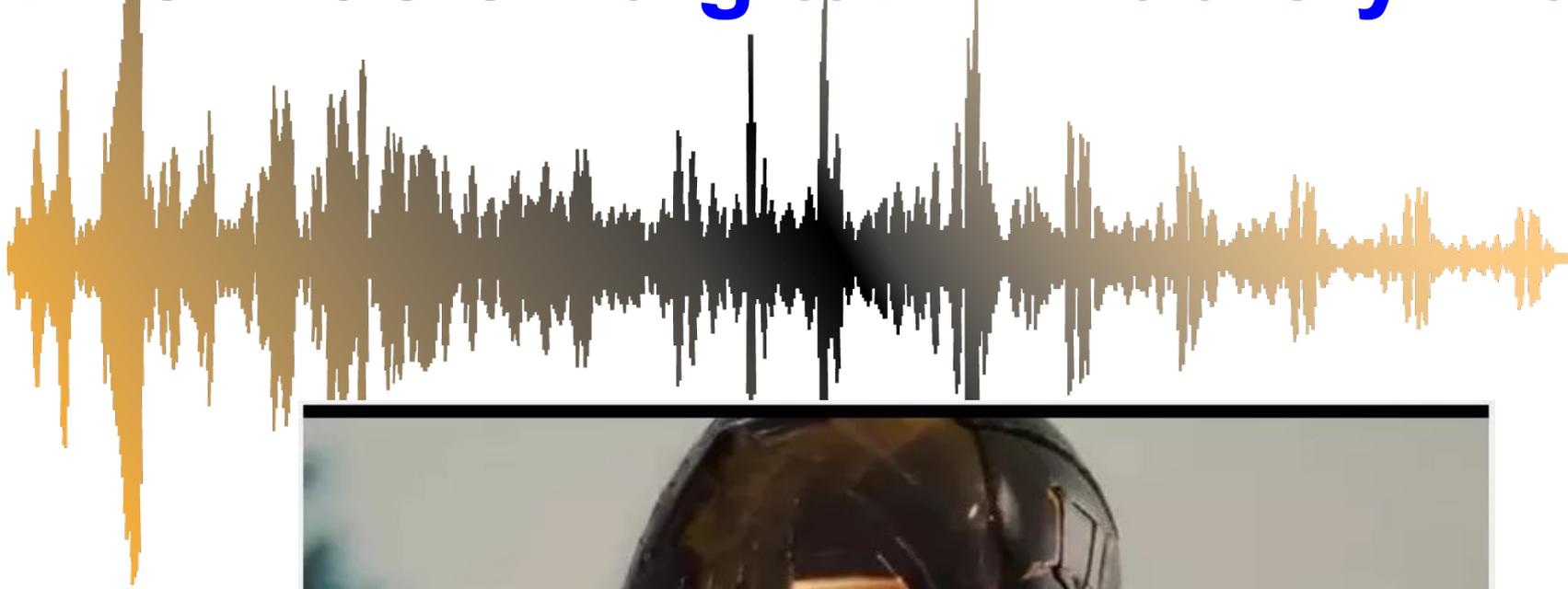
Subject

-----BEGIN PGP MESSAGE-----
Version: OpenPGP.js v.1.20130712
Comment: <http://openpgpjs.org>

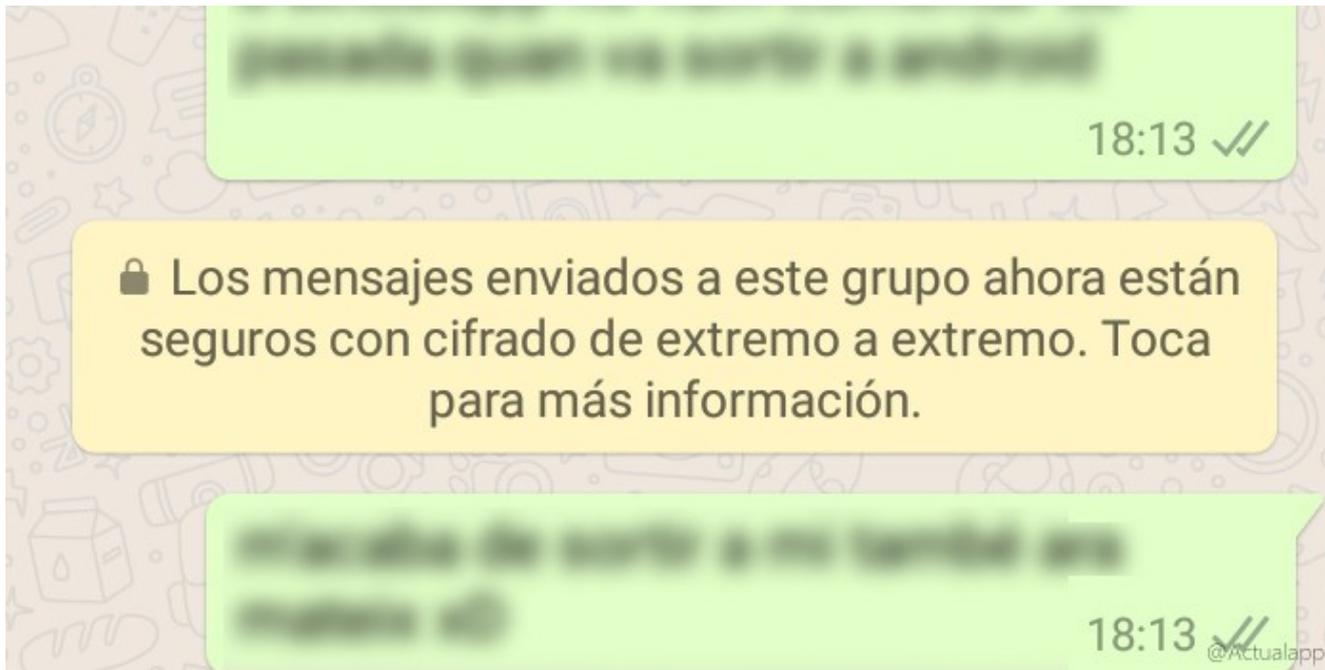


wYwDzoryw22LIbsBBACgY0920+TBs9wr3imXPonZ/sul3z1nTwMVfqZl7BHb
49KlbqKwePIQnLZeu2ckfybJsC4enHQTzn8ZTJlvIETHf/o0plEF87jf82MI
mRDekE1fcdj5consy2VFAsJahTHOOFBxoLk72m7iuPK0mf0SgN8hX+R53MzJ
1RC+tBoafcHATAOp8CsqJMp8SgEIALrtDdwuq0eCqjw7Jv3cAf7OBVxC/74F
rCDWI1FC6TI3IBVXlrZ2182/GGND+yvWXyZsP5vBAdVlkv+mbApsvvCWYc5p
oybkxpuCcT1IWBQRAkZtdePNXTkNH6yHiagkOdYLoGsTI9ThZI4jHtY8DJ2u
+iFO3tsAY3UD0zeLploLmmvwmOFFEb6DUUVdU1Uglko4uMxG6qAvc6dTLGmPQ
Y/VKjNw5TgOFeKFAAgSgmkr7eAcmnD7cn2p9fKJpGq50BWgi6cOoMxZth0Z
8CIP7n2CbHQFQUvK1F/alrLhDQ7NCInCosOYMwF2VmPYP/AU1DH/8khdGGCe
jusKjBS/nnvSrQFdViFOGgcEFvTTzrCb1mKR1ImCwB2DIUreclmXpGzs9aWr
uHlqjuq2BjagW9qOD3+V5KIVt86CaUjlAAISnqW08nRZVWe/dktDhE7fvfUx
VBZfX+CFLok17rNxfKdnrB/PbW4zqL5w1erO9A6QYMXOqY3BLbYjA3eyzs9
JPMvUsAeSjK8CEct7uV8qv348CisGZT0nfQlXuE4H+0qnMS1sUKHPF5m6X7b
Yflg
=D7x6
-----END PGP MESSAGE-----

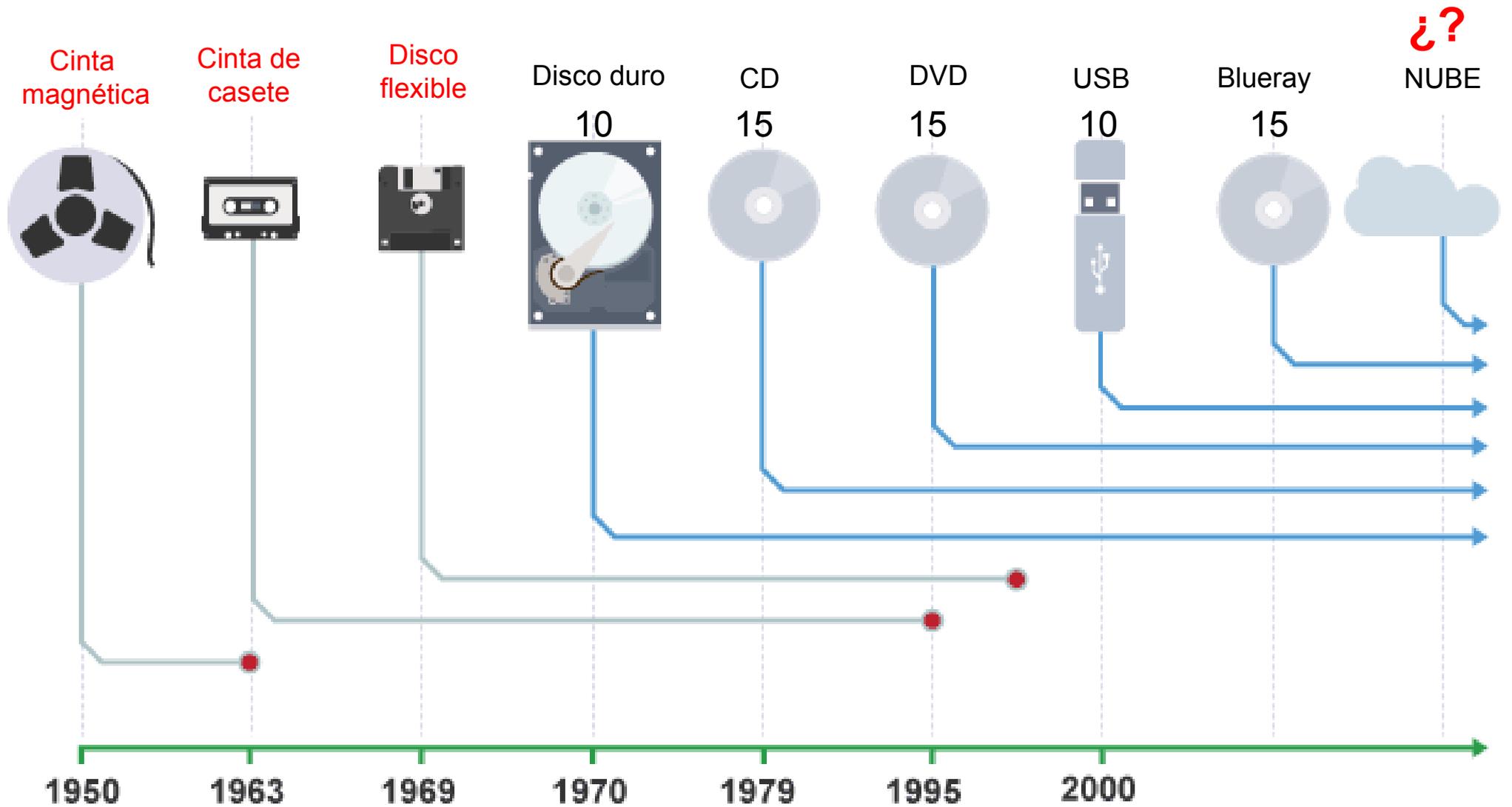
Información digital → Audio y Video



Información digital → Email



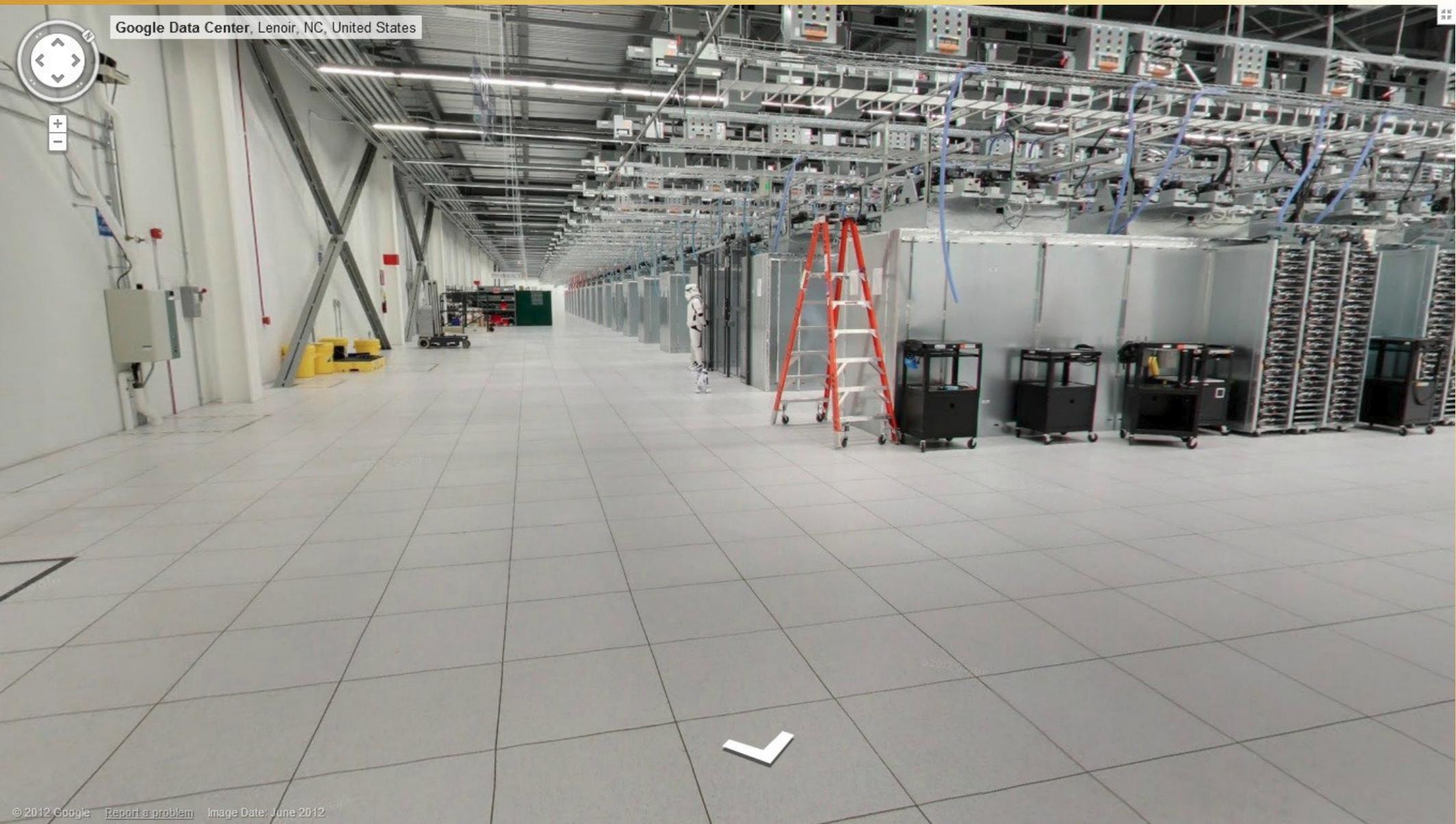
¿Información?



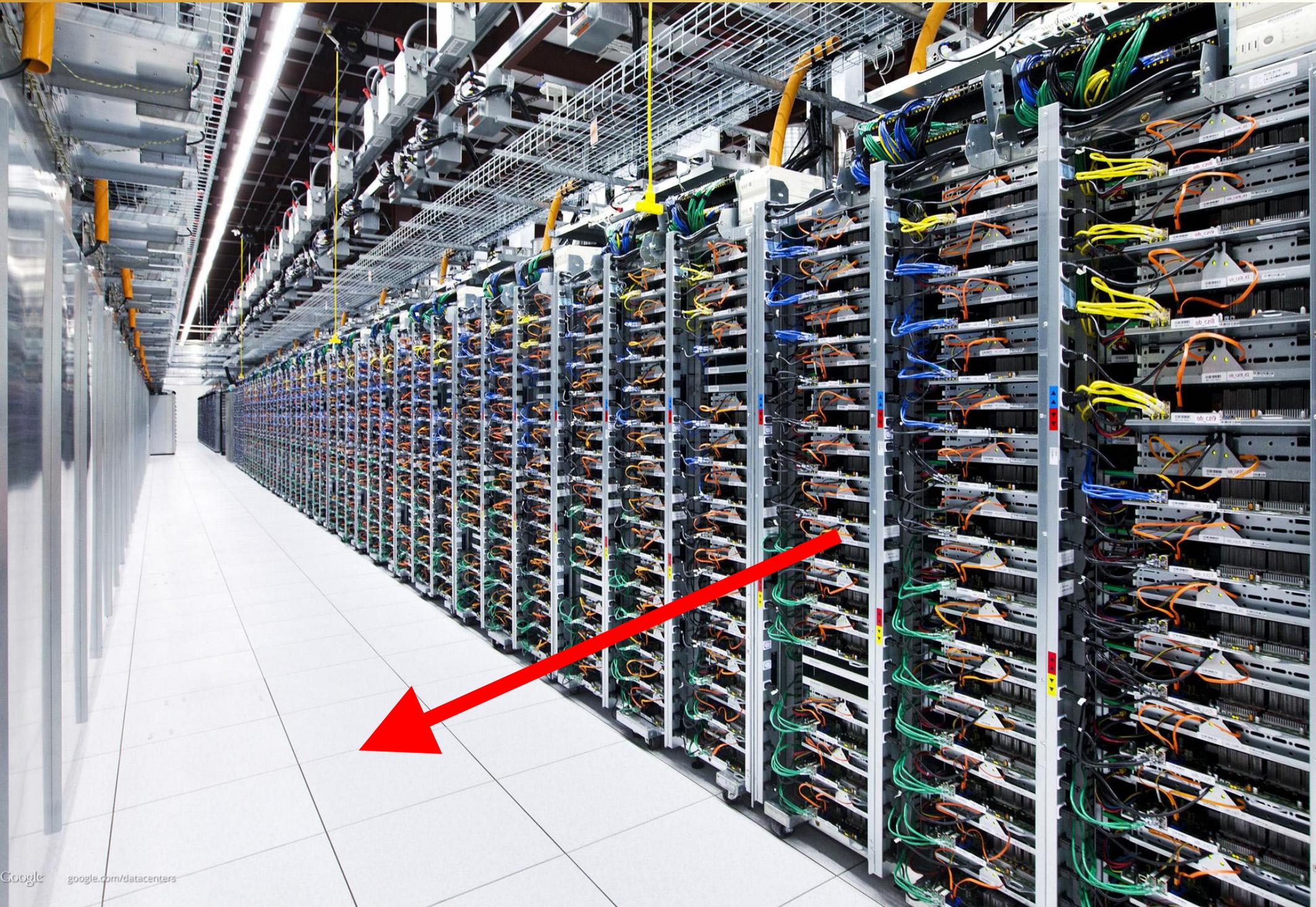
La NUBE



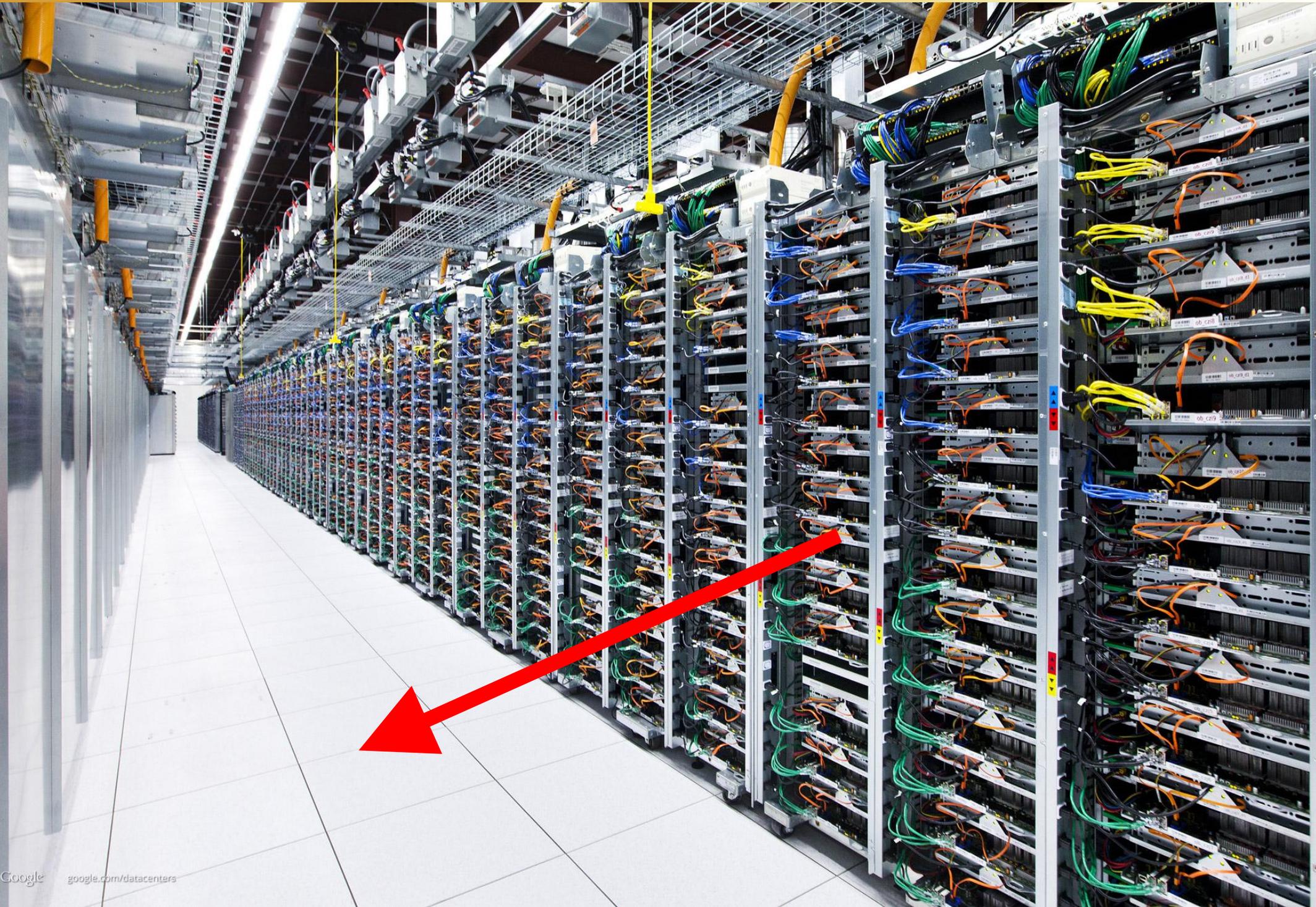
La NUBE: Datacenter



La NUBE: saquemos un equipo



La NUBE: saquemos un equipo

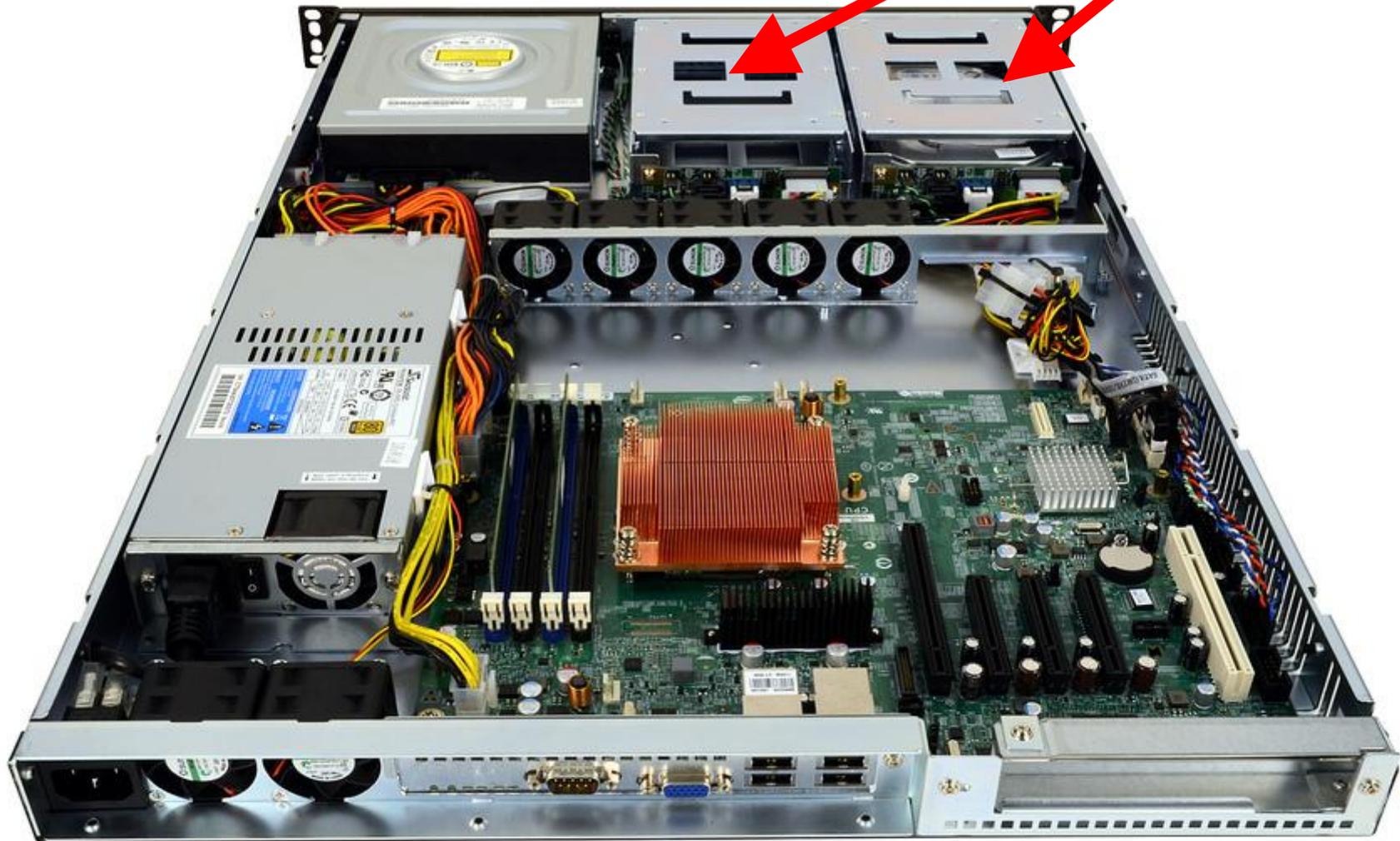


La NUBE: abramos el equipo



La NUBE: interior del ordenador

Disco duro



1 disco duro = 10 años = 3 TBytes



10 años



¡Sincronicemos todo el tiempo!



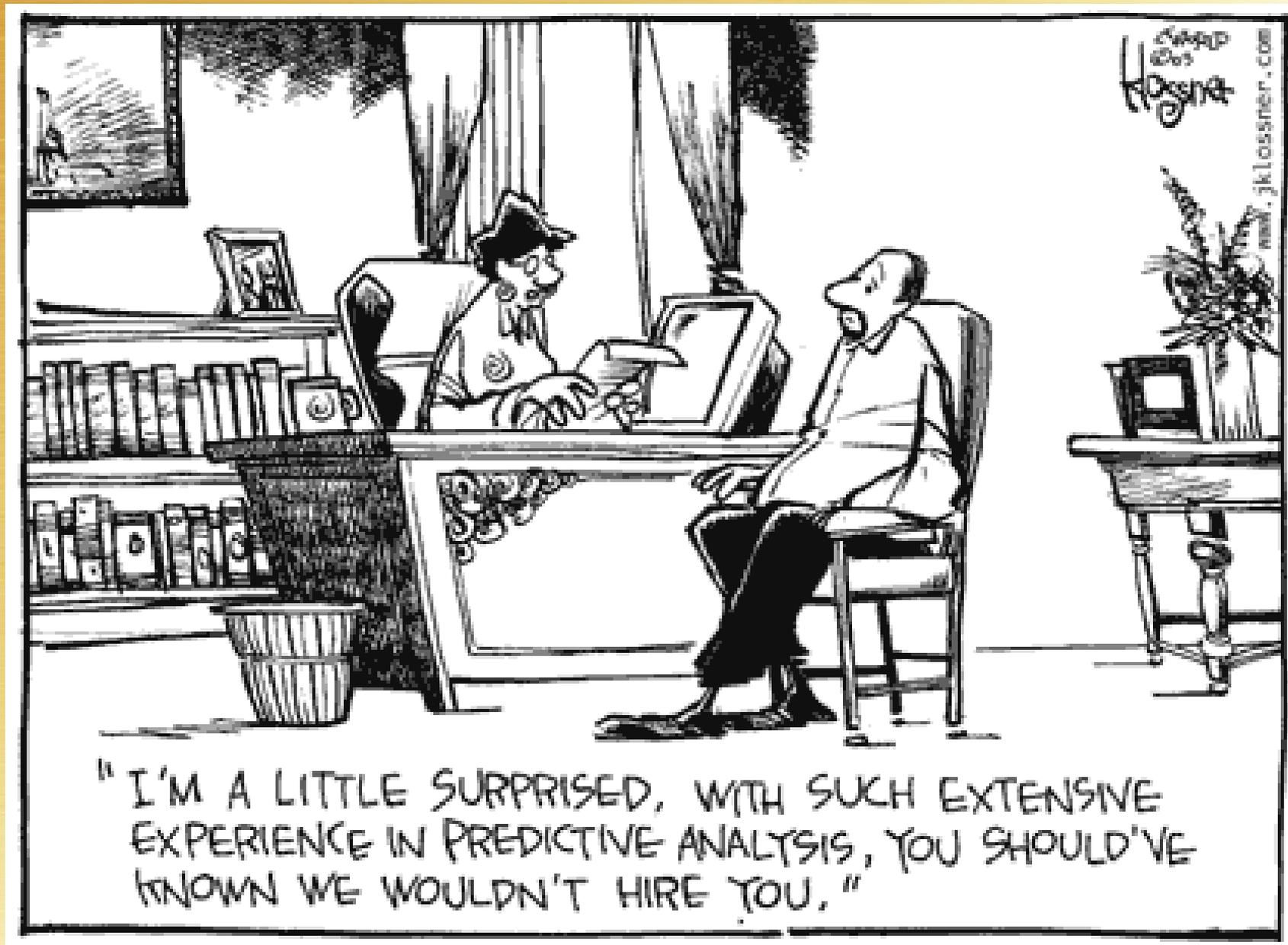
¿Cómo pinta todo esto?



Juanmi Taboada

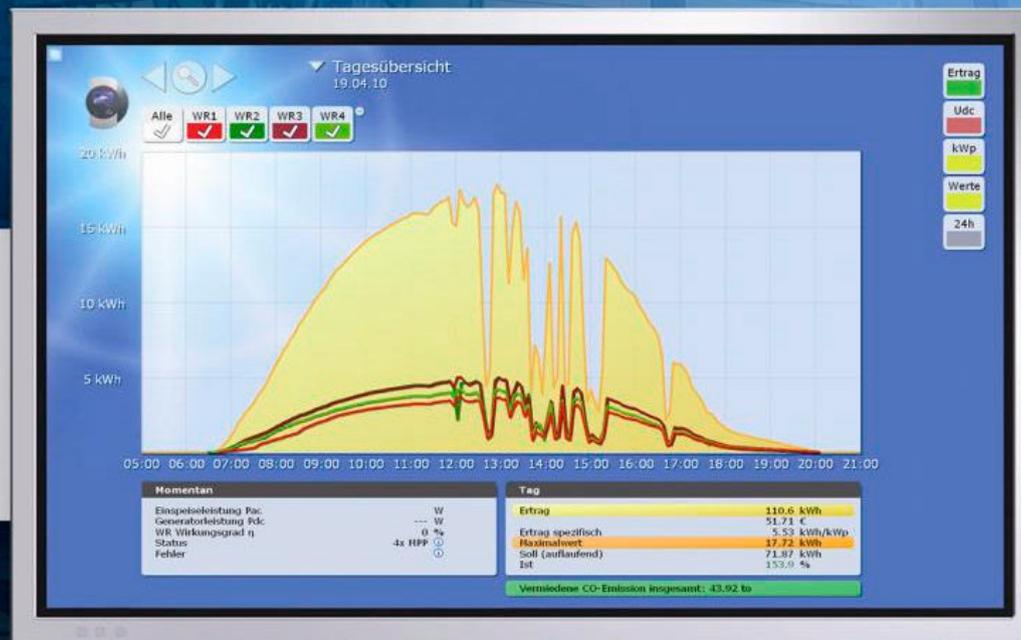
Big Data

Una dosis de realidad





Monitorización



Plant example 2

Zeit	Typ	Wärmeleistung (kWh)	Udc (€)	kWp (kWh/kWp)	Werte (kWh)	24h (kWh)
18.01.2010 22:30:16	Verbindung	18000	1000	1000	1000	1000
17.01.2010 22:30:36	Verbindung	18000	1000	1000	1000	1000
16.01.2010 22:30:16	Verbindung	18000	1000	1000	1000	1000
15.01.2010 22:30:36	Verbindung	18000	1000	1000	1000	1000

Alle Meldungen des Typs quittieren

Plant example 3

Top!



Likingdoy

The monitoring company

Software de adquisición masiva de datos:

- 1 dato (Fecha+Valor) cada minuto por se al
- 1000 se ales por dispositivo
- 20 dispositivos por nodo
- 40 nodos por cliente

800.000 registros por minuto
(8.000.000 tomas por minuto)

- 48M por hora
- 1.152M por día

0,4 Billones por año y cliente

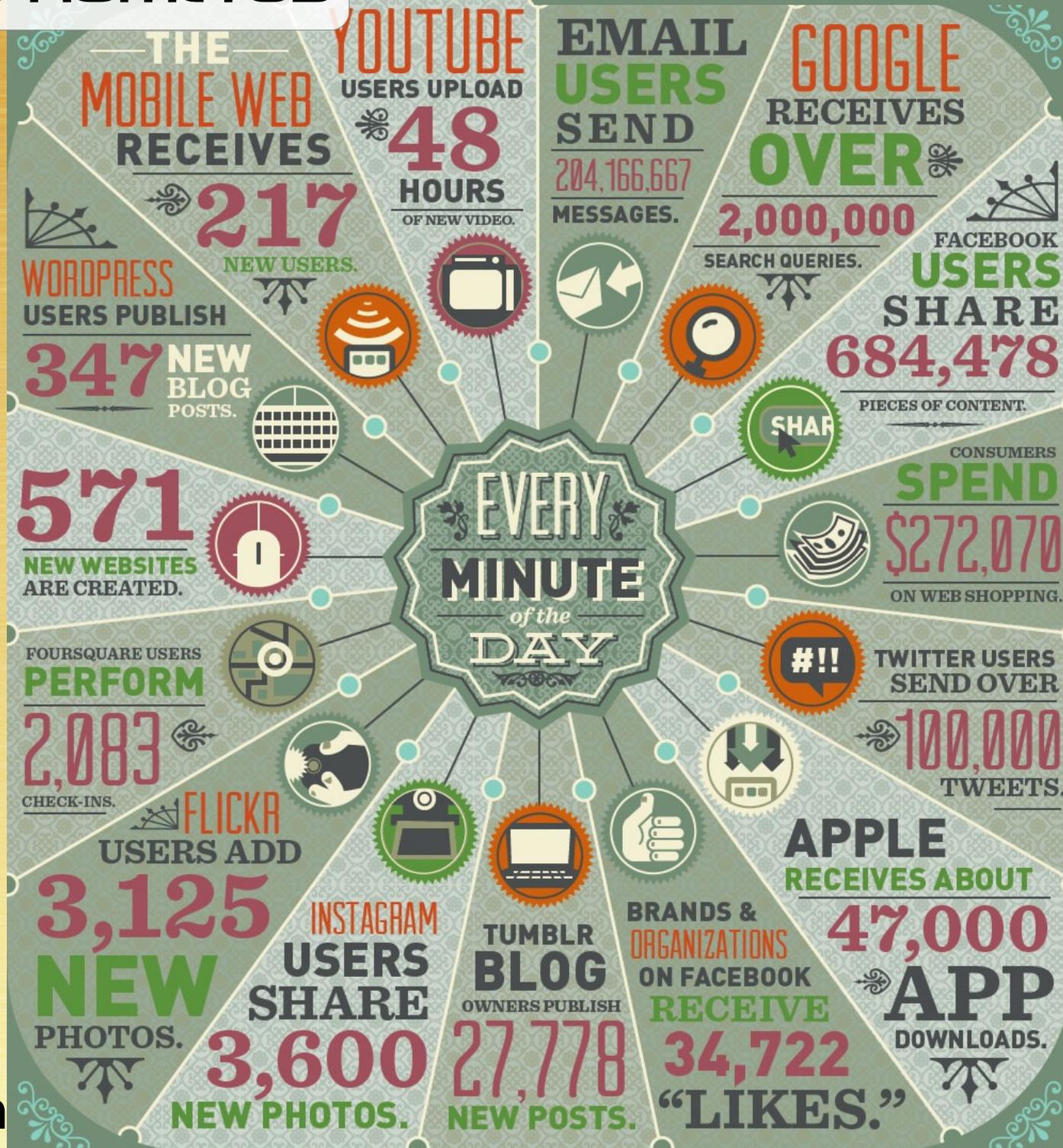


Grandes números

SOURCES: [HTTP://NEWS.INVESTORS.COM/](http://news.investors.com/), [ROYAL.PINGDOM.COM](http://royal.pingdom.com), [BLOG.GROVO.COM](http://blog.grovo.com), [BLOG.HUBSPOT.COM](http://blog.hubspot.com), [SIMPLYZESTY.COM](http://simplyzesty.com), [PCWORLD.COM](http://pcworld.com), [BIZTECHMAGAZINE.COM](http://biztechmagazine.com), [DIGBY.COM](http://digby.com)

2012

2.1 Billones de usuarios en Internet



Juanm Data

Grandes números

2014

De 2.1
a 2.4
billones

2012 - 2.1
2014 - 2.4

PINTEREST
USERS PIN

3,472
images.

YOUTUBE
USERS UPLOAD

72 HRS.
OF NEW
VIDEO.

EMAIL
USERS SEND
204,000,000
MESSAGES.

Google

RECEIVES OVER
4,000,000
SEARCH
QUERIES.

FACEBOOK
USERS SHARE

2,460,000
PIECES OF CONTENT.

VINE
USERS

SHARE
8,333
VIDEOS.

SKYPE
USERS
CONNECT FOR
23,300 HOURS.

EVERY
MINUTE
OF THE
DAY

TINDER
USERS SWIPE
416,667
TIMES.

YELP USERS
POST
26,380
REVIEWS.

WHATSAPP
— USERS SHARE —
347,222
PHOTOS.

APPLE USERS
DOWNLOAD

48,000
apps.

PANDORA
USERS LISTEN TO

61,141
HOURS OF
music.

AMAZON
MAKES
\$83,000
IN ONLINE SALES.

TWITTER USERS

TWEET

INSTAGRAM
USERS »

POST

216,000
NEW PHOTOS.

277,000
TIMES.

SOURCES:

BITS.BLOGS.NYTIMES.COM, INTEL.COM, APPLE.COM, TIME.COM, DAILYMAIL.CO.UK, SKYPE.COM, STATISTICBRAIN.COM



Juanmi

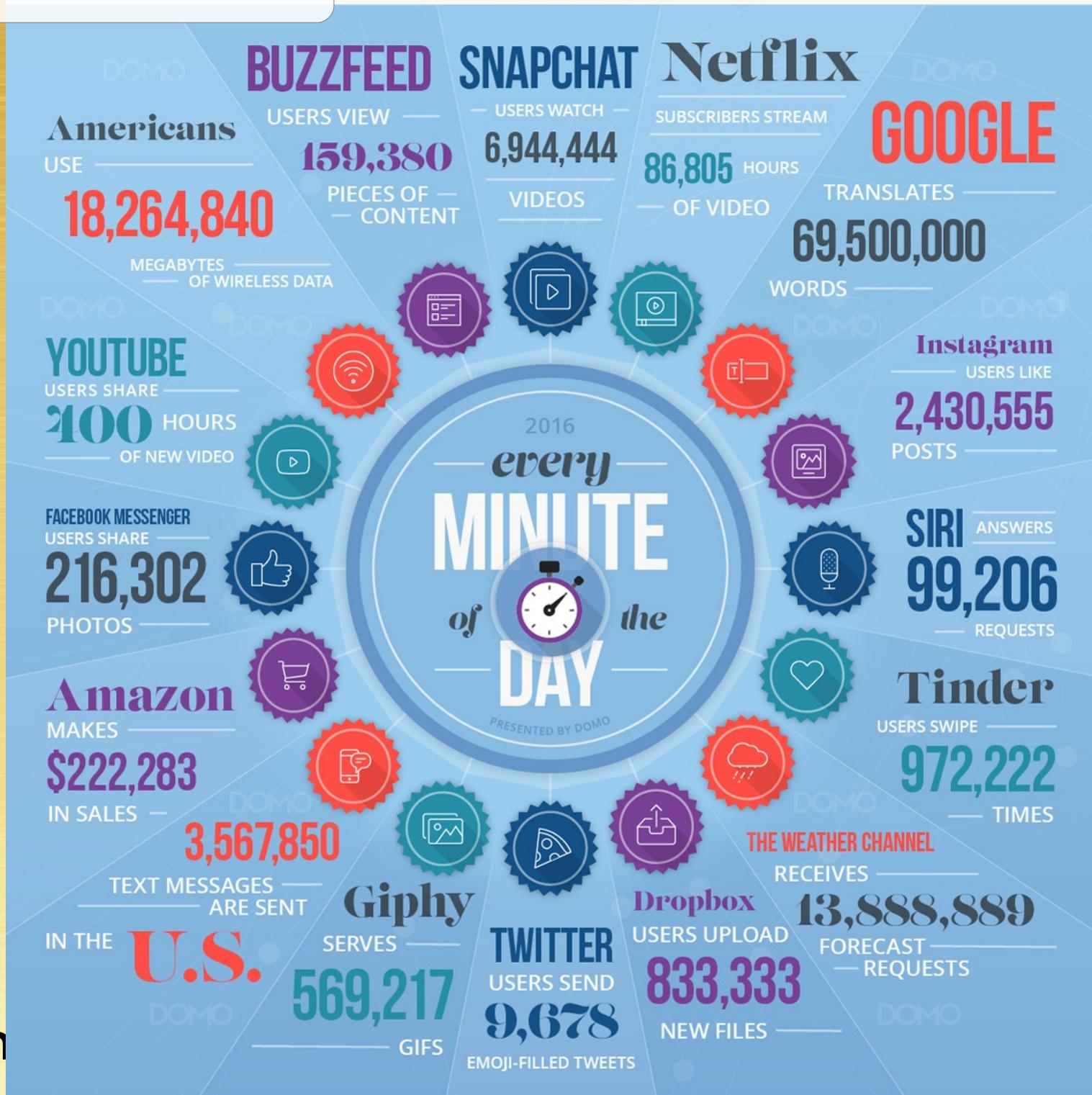
Data

Grandes números

2016

De 2.4
a 3.4
billones

2012 - 2.1
2014 - 2.4
2016 - 3.4

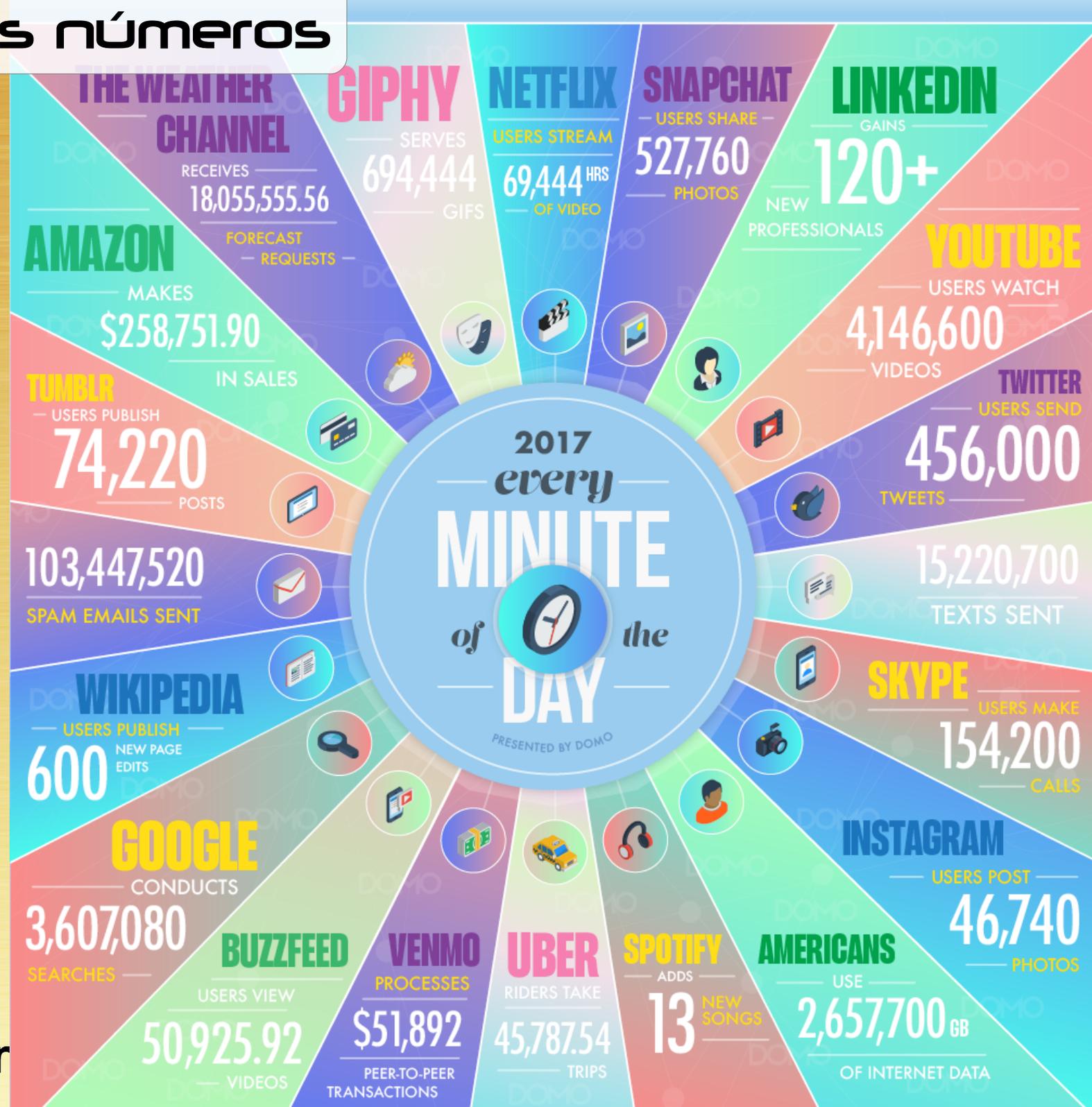


Grandes números

2017

De 3.4 a 3.7 billones

- 2012 - 2.1
- 2014 - 2.4
- 2016 - 3.4
- 2017 - 3.7

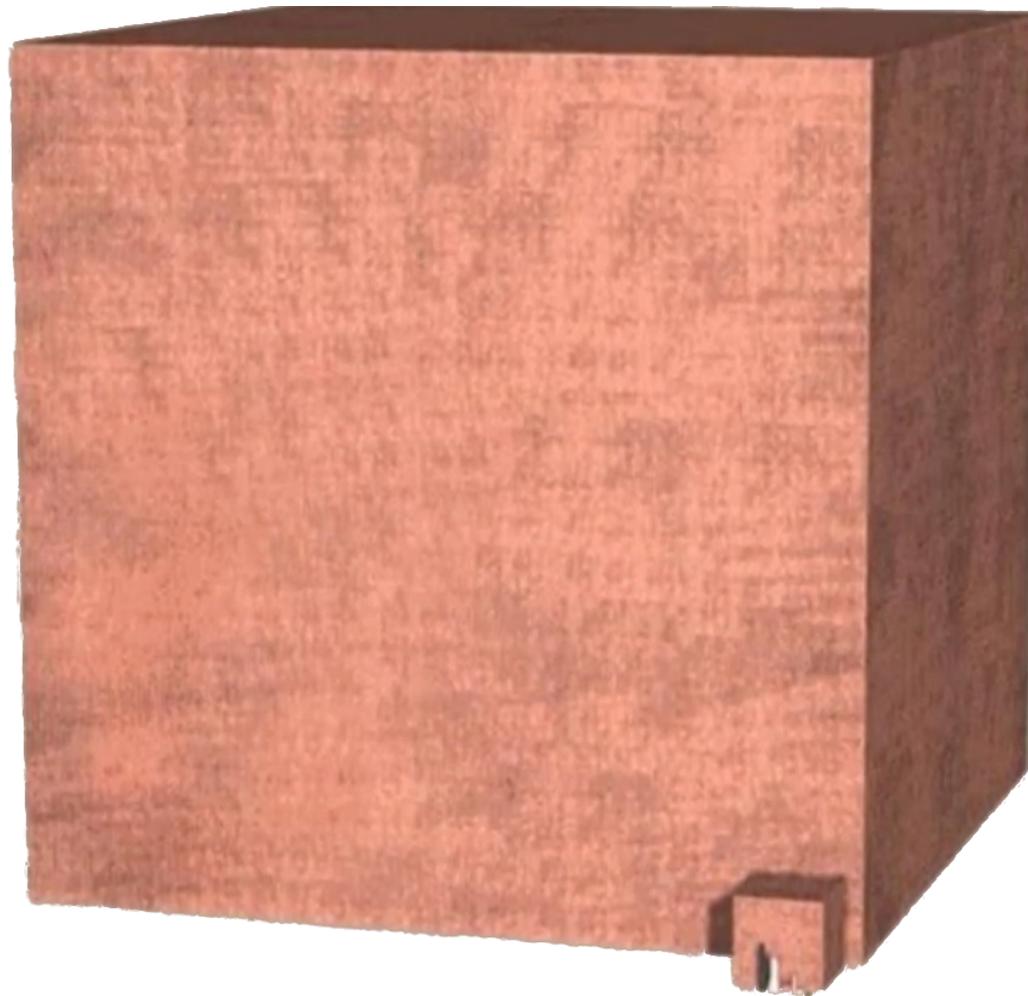


¿Cuanto es un quintillón?

- Una única **gota de agua** contiene:
1.7 quintillones de moléculas de agua.
- La distancia de la **Via Láctea hasta Andrómeda** es de:
2 millones de años luz
18,87 quintillones de kilómetros
11,73 quintillon miles
- **La tierra** completa contiene unos:
1.234 quintillones de litros de agua
326 quintillon gallons of water
- Si cortamos **la tierra por la mitad**, la sección tendría un área aproximada de:
1.275 quintillones de centímetros cuadrados
- ¿Cuanto es un quintillón de céntimos o peniques de dólar?



¿Cuanto es un quintillón... de peniques?



1.000.067.088.384.000.000 peniques

1 quintillón, 67 trillones, 88 billones, 384 millones de peniques

Un cubo de **8,32 kilómetros** de lado



- La gestión de datos crece de momento EXPONENCIALMENTE
- Seagate dice: **2016 año del Zettabyte (16ZB) y 163 ZB para 2025**
- 1 ZB equivale al espacio que ocupa 2 billones de años de música
- 1 ZB = 1024 exabytes = casi 1.1 trillones de Gbytes
- 1 zB = 1 sextillón de bytes



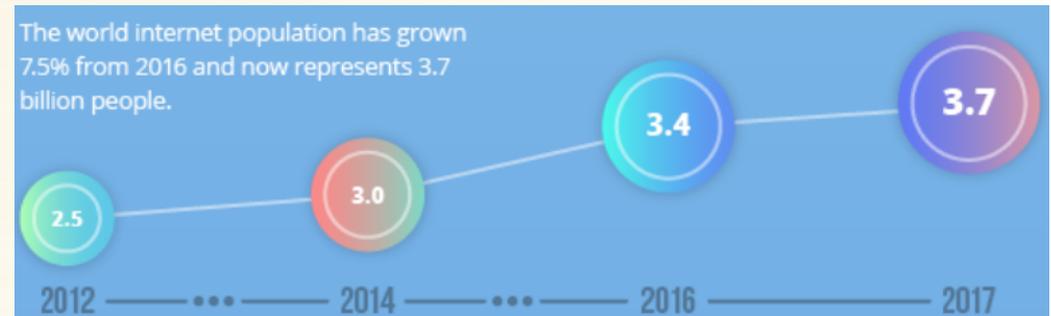
Usuarios de Internet:

2012 – 2.1 Billones

2014 – 2.4 Billones

2016 – 3.4 Billones

2017 – 3.7 Billones



Un usuario genera al mes:

en 2011: 11,5 Gbytes

en 2016: 32,3 Gbytes

Hoy: 86% del tráfico son videos

Toda la historia de películas juntas cruzan Internet **cada 3 minutos**

A lo largo de 1 año hablamos de 1 trillón de Gigabytes o ...

1 Zetabyte

Eric Brewer (2000)

- C: Consistency** → **Consistencia**
- A: Availability** → **Disponibilidad**
- P: Partition tolerance** → **Tolerancia al particionado**

Sólo puedes llegar a 2 de las 3

Es imposible para un sistema de cómputo distribuido garantizar simultáneamente la **consistencia**, la **disponibilidad** y ser tolerante al **particionado** de los datos (separación y distribución).

Teorema de CAP

Lo que la mayoría **piensa** que tiene

Siempre se puede leer y escribir

A

C

P

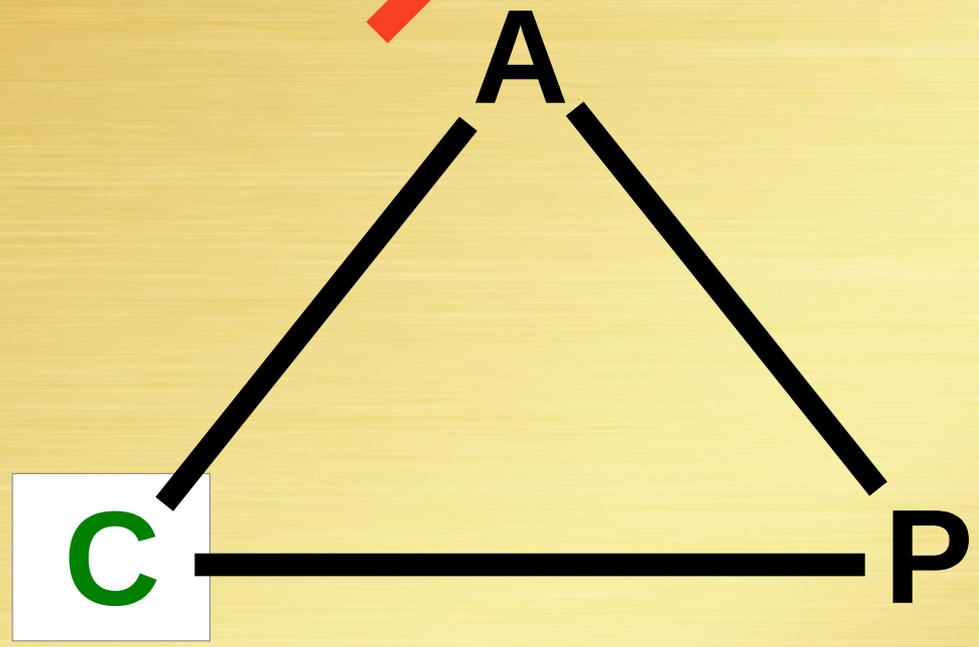
Todos los clientes ven siempre lo mismo

Funcionará incluso cuando haya errores en el sistema

Teorema de CAP

Lo que la mayoría de **verdad** tiene

~~Siempre se puede leer y escribir~~



Todos los clientes ven siempre lo mismo

~~Funcionará incluso cuando haya errores en el sistema~~

Lo cierto es que todos buscamos la **disponibilidad (A)**

Pero ... ¡¡¡ tenemos que elegir entre... !!!

Escalabilidad (P)

y

Consistencia (C)

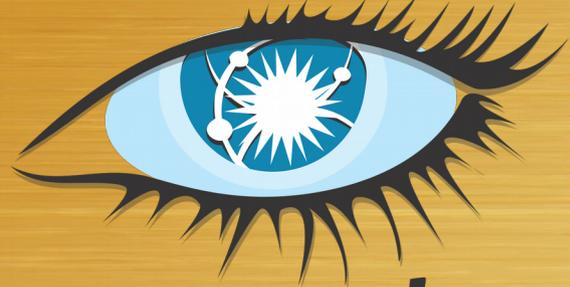
¿Qué es la consistencia eventual?

Que ... eventualmente será consistente

Podemos introducir un dato y que no esté disponible inmediatamente después

Convergencia natural a la consistencia

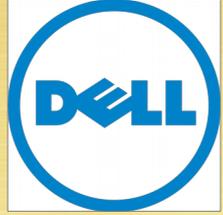




cassandra

¿Quién usa Cassandra?

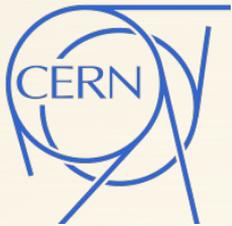
Fuente: <http://planetcassandra.org/companies/>



Juanmi Taboada

BIG DATA

¿Quién usa Big Data?



Al finalizar la última iteración del LHC en el que se descubrió el “**Bosón de Higgs**”, el CERN almacenaba más de **100Pbytes**

El LHC del CERN arrancó en **Abril de 2015** en busca de la **Supersimetría** (capaz de producir 1Pbytes/segundo)

1Petabyte = 1.000 Terabytes = 1 Millón de Gigabytes

NETFLIX

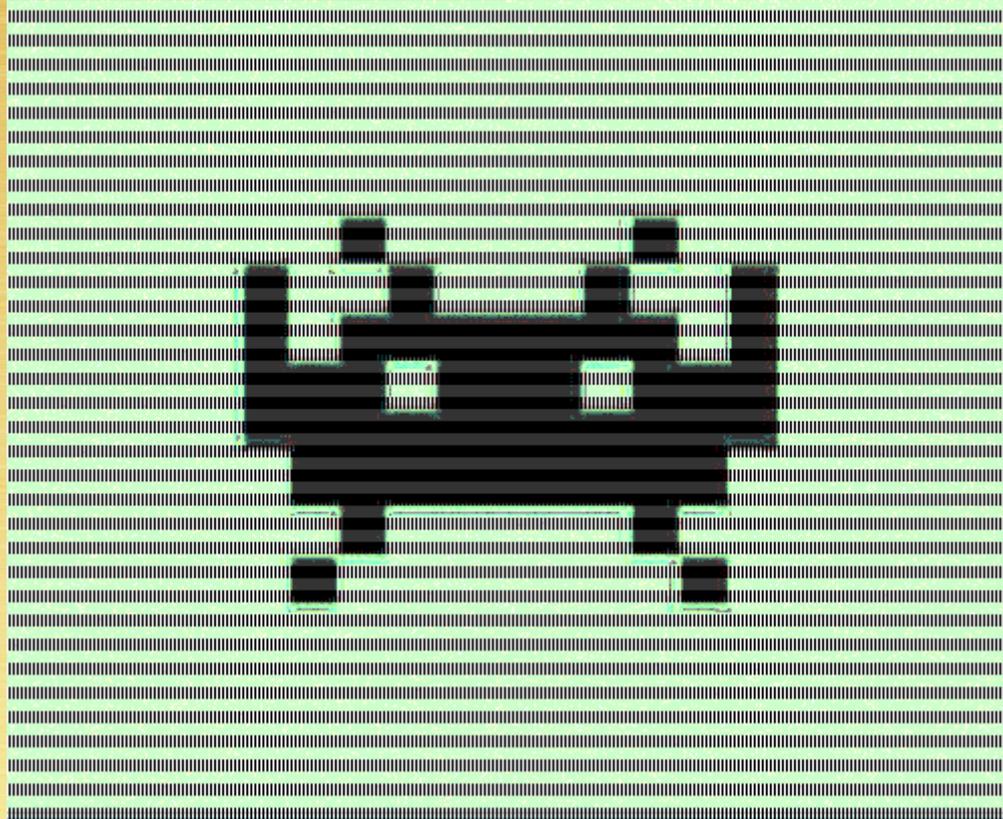
Ejecuta 235 clusters separados, con un total de 7.000 nodos.
1 Millón de escrituras por segundo (factor 3)



Ecosistema para desarrolladores de juegos que teniendo problemas con MongoDB migraron a Cassandra.



¿Dudas?



Muchas Gracias

Thank you - Dziękuję



Juanmi Taboada

<http://www.juanmitaboada.com>



@juanmitaboada

<https://es.linkedin.com/in/juanmitaboada>



Juanmi Taboada

Big Data