

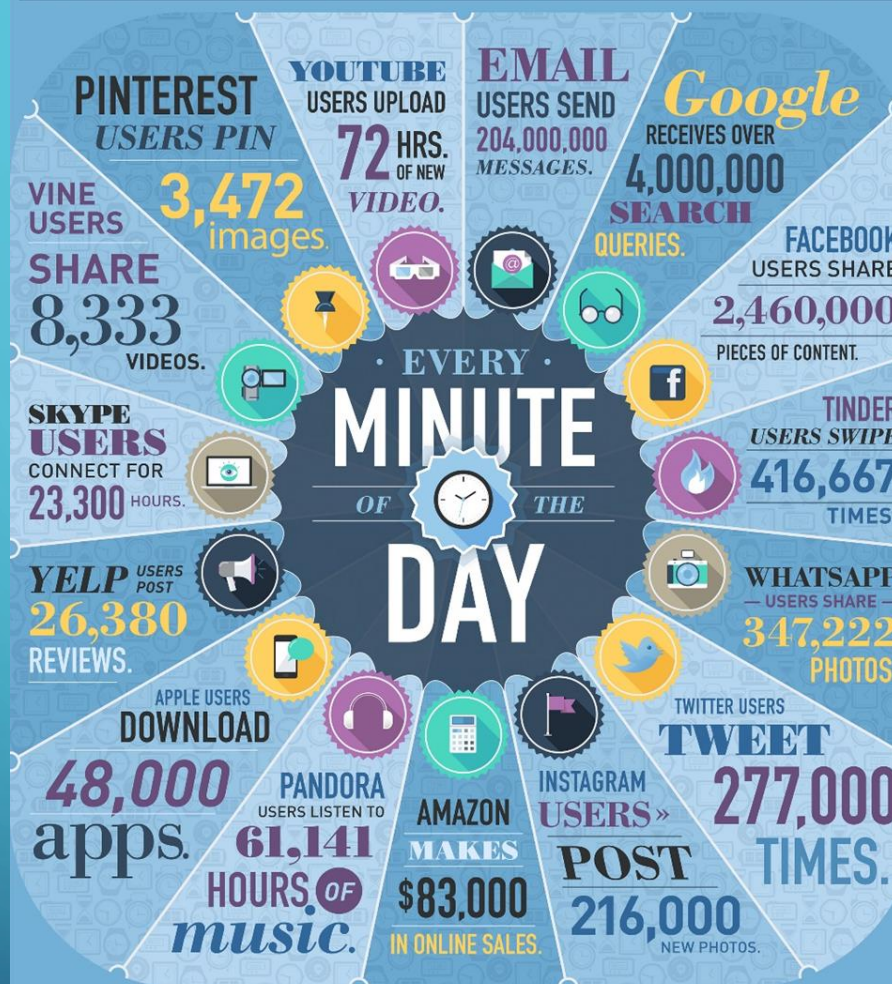
BIG DATA



DATA NEVER SLEEPS 2.0

How Much Data is Generated **Every Minute?**

Data is being created every minute of every day without us even noticing it. Given how much information is flooding around these days, it's tempting to talk about big data only in terms of size. Big data describes the massive avalanche of digital activity pulsating through cables and airwaves, but it also describes all the things we were never able to measure before. With every status we share, every article we read or every photo we upload, we are creating a digital trail that tells a story. Below, we explore how much data is generated in one minute.



THE GLOBAL INTERNET POPULATION GREW **14.3% FROM 2011 - 2013** AND NOW REPRESENTS **2.4 BILLION PEOPLE.**

With each click, share and like, the world's data pool is expanding faster than we can comprehend. Businesses today are paying attention to scores of data sources to make crucial decisions about the future. The team at Domo can help your business make sense of this endless stream of data by providing executives with all their critical information in one intuitive platform. Domo delivers the insights you need to transform the way you run your business. Learn more at www.domo.com.

SOURCES:

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



DATA NEVER SLEEPS 8.0

How much data is generated **every minute?**

In 2020, the world changed fundamentally—and so did the data that makes the world go round. As COVID-19 swept the globe, nearly every aspect of life—from work to working out—moved online, and people depended more and more on apps and the internet to socialize, educate and entertain ourselves. Before quarantine, just 15% of Americans worked from home. Now over half do. And that's not the only big shift. In our 8th edition of Data Never Sleeps, we bring you the latest stats on how much data is being created in every digital minute—a trend that shows no sign of stopping.



The world's internet population is growing significantly year over year. As of April 2020, the internet reaches 59% of the world's population and now represents 4.57 billion people — a 6% increase from January 2019.



GLOBAL INTERNET POPULATION GROWTH 2014-2020 (IN BILLIONS)

As the world changes, businesses need to change with the times—and that requires data. Every click, swipe, share or like tells you something about your customers and what they want, and Domo is here to help your business make sense of all of it. Domo gives you the power to make data-driven decisions at any moment, on any device, so you can make smart choices in a rapidly changing world.

Learn more at domo.com

SOURCE: STATISTA, VALUE CAPITALIST, BUSINESS INSIDER, GAMESPOT, TECHCRUNCH, COMSCORE AGENCY, DOORDASH, BUSINESS OF APPS, NEW YORK TIMES, MUSIC BUSINESS WORLDWIDE, INC., THE VERGE, INC., HOOVER/STRAT, QUANTIL STUDIOS, REDDIT, USER, AMAZON, VISA



FUENTE [HTTPS://WWW.DOMO.COM/LEARN/DATA-NEVER-SLEEPS-8](https://www.domo.com/learn/data-never-sleeps-8)

Big Data:

Los **macrodatos**, también llamados **datos masivos**, **inteligencia de datos**, **datos a gran escala** o ***big data***, es un término que hace referencia a conjuntos de datos tan grandes y complejos que precisan de aplicaciones informáticas no tradicionales de procesamiento de datos para tratarlos adecuadamente.

- En general se enfoca en el conjunto de herramientas que permiten revelar relaciones y dependencias, o para realizar predicciones de resultados y comportamientos.

THE 5V'S: TURNING BIG DATA INTO VALUE

With the datafication comes big data, which is often described using the four Vs:

THE DATAFICATION OF OUR WORLD:



ACTIVITY DATA

Music players, eReaders and smart phones collect data on how we use them; web browsers collect information on what we search for; credit card companies collect data on where we shop; and shops collect data on what we buy.



CONVERSATION DATA

Our conversations are being captured - From emails to all the conversations we have on social media sites like Facebook or Twitter as well as our phone conversations are now digitally recorded.



PHOTO AND VIDEO IMAGE DATA

All the pictures and videos we take on our smart phones and digital cameras - we upload and share millions of them on social media sites every second.



SENSOR DATA

We are surrounded by sensors that collect and share data - devices like our smart phones use sensors to track our location, the speed and direction at which we are traveling, read our fingerprints, detect how light it is outside, etc.

VOLUME

...refers to the vast amounts of data generated every second - Today, we create the same amount of data in a single minute, that was created from the beginning of time until the year 2000.

VELOCITY

...refers to the speed at which new data is generated and the speed at which data moves around - Today, we perform millions of internet searches every second, social media messages can go viral in minutes, and credit card transactions are checked in real-time.

VARIETY

...refers to the different types of data we can now use - Today, we don't have to rely on nicely structured data, we can now collect and analyse text, images, video, voice, location data, and much more.

VERACITY

...refers to the messiness or trustworthiness of the data - Today, quality and accuracy of data are less controllable (hash tags, abbreviations, typos and colloquial speech) but technology now allows us to deal with it.

VALUE

...the final V refers to the need to turn our data into value - Today, big data is used to better understand and target customers, understand and optimize business processes, and improve health care, security and law enforcement. But the possible applications of big data are endless!

ANALYZING BIG DATA:

TEXT ANALYTICS

SENTIMENT ANALYTICS

FACE RECOGNITION

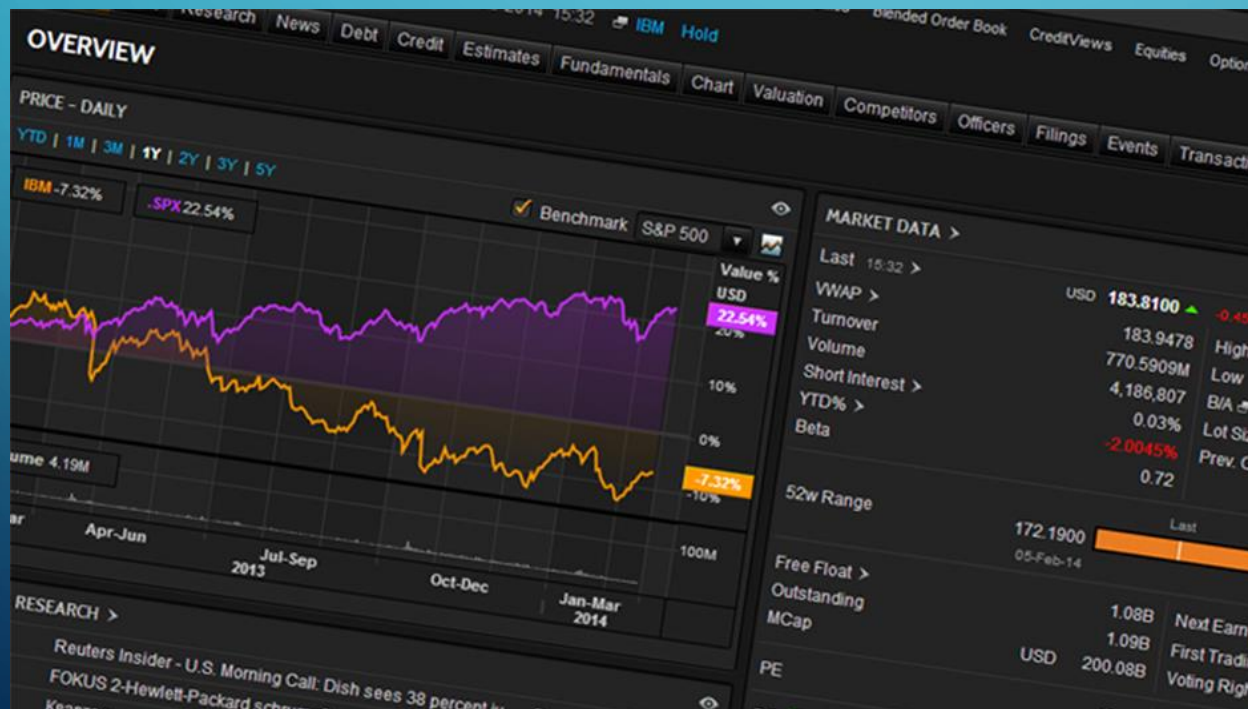
VOICE ANALYTICS

MOVEMENT ANALYTICS

Volumen

Estamos viviendo en una época en la que la información nos *rodea por completo*

Mercados financieros:

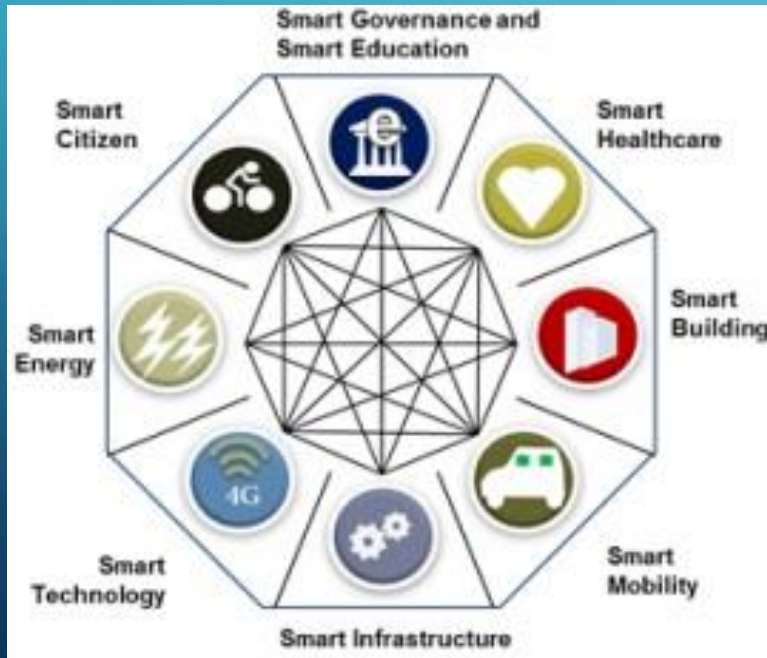


En los últimos 20 años se produjeron unos 20PB de información sobre millones de instrumentos financieros en más de 400 mercados, con datos tomados *cada μ s*.

Otros Problemas: *Fusión*

Estamos viviendo en una época en la que la información nos *rodea por completo*

Ciudades inteligentes y gobierno digital

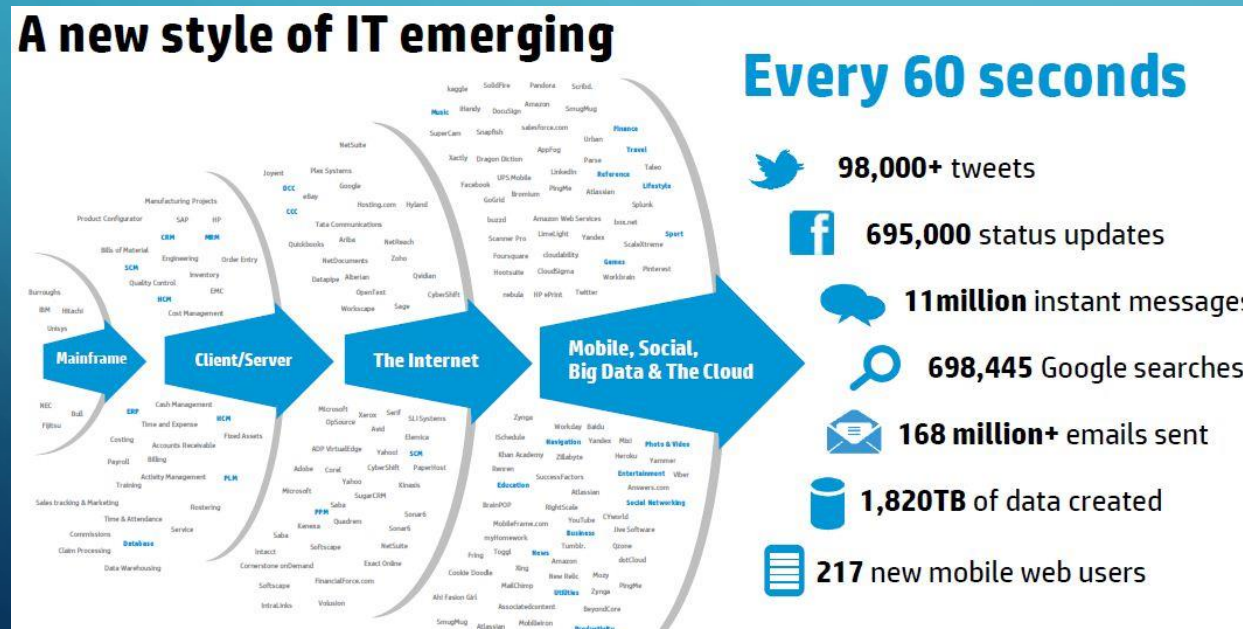


Estas aplicaciones requieren soluciones al problema de la *fusión* de datos para ser efectivos: tránsito (aéreo, trenes, autos, camiones), horarios de transporte público, cámaras, sensores ambientales, estacionamiento, manifestaciones, ...

Otros Problemas: *Tipos de datos nuevos*

Estamos viviendo en una época en la que la información nos *rodea por completo*

Redes sociales



Otros Problemas: *Tipos de datos nuevos*



Otros Problemas: *Tipos de datos nuevos*

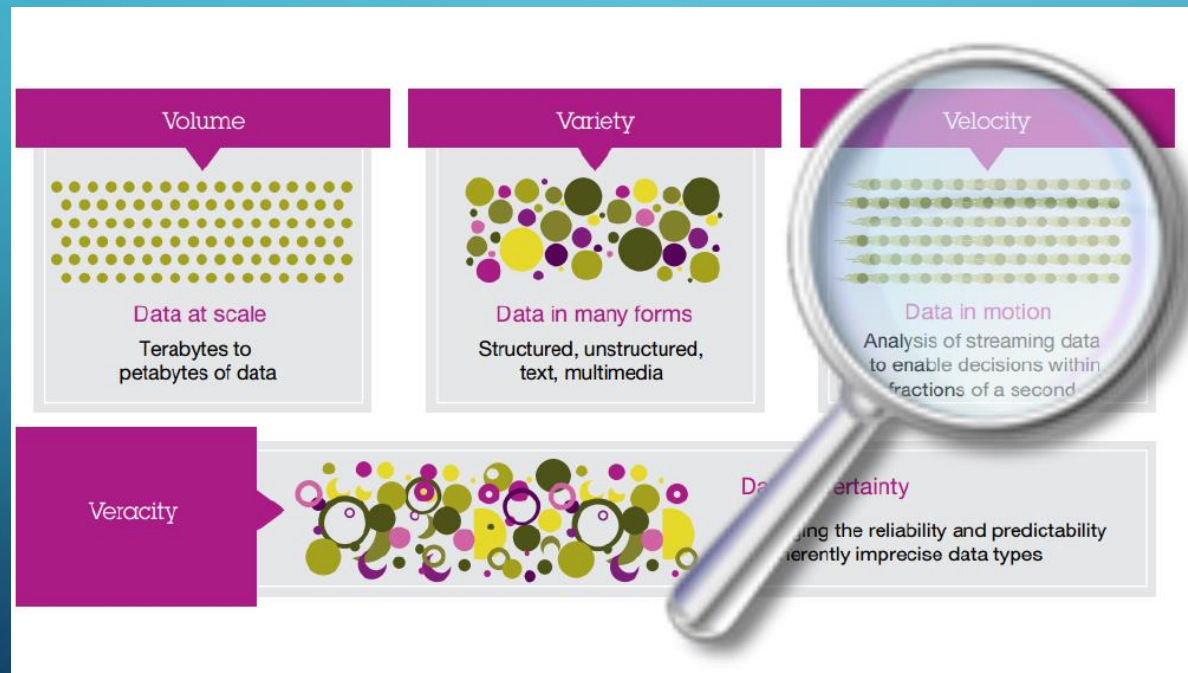


Una “tormenta perfecta” de desafíos:

- 1) cantidades *masivas* de información producida constantemente;
- 2) altas *tasas* de crecimiento; y
- 3) nuevos *tipos* de datos.

Las 4 Vs del Big Data: *Velocidad*

Este aluvión de información debe procesarse de manera eficiente y efectiva para aprovechar su contenido.



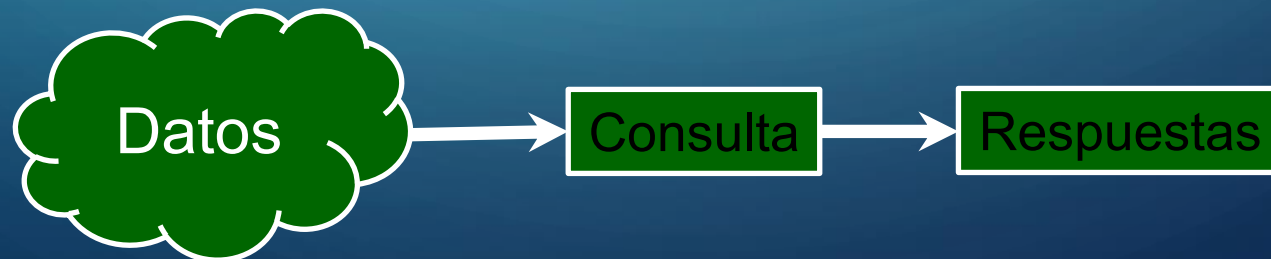
Fuente: [DCM15]

Velocidad: ¿Adiós a los DBMS?

- El modelo tradicional de las bases de datos relacionales sigue el siguiente esquema:



- Sin embargo, dado que *no podemos almacenar* todos los datos que se generan, se debe seguir otro esquema:



Velocidad: ¿Adiós a los DBMS?

- Ejemplo: compra/venta (*trading*) de acciones a alta frecuencia:
 - Las operaciones se llevan a cabo en tan sólo 1ms.
 - La **proximidad física** a los centros de cómputo supone una ventaja sobre los competidores;
 - No hay estadísticas confiables, pero se estima que entre el 60% y 70% de las operaciones de EE.UU. se llevan a cabo a alta frecuencia (alrededor del 40% en Europa).
- La detección de **fraude** en este entorno es un desafío:
 - **Fusión** de flujos de información para detectar anomalías.
 - Debe hacerse en tiempo “casi real” (*near real-time*) para mantener la confianza de los consumidores.

Velocidad: ¿Adiós a los DBMS?

Ya hace más de **10 años** que se planteó la pregunta, y las respuestas han sido contundentes:

- Se pasó de procesar datos de un sólo negocio a un **conjunto**, con **requerimientos heterogéneos**.
- Nuevas características: “*shared nothing*” y alta disponibilidad.
- “*No knobs*”: los DBMS son **altamente configurables** porque antes el personal era barato y el hardware costoso; hoy es a la **inversa**.
- *Multi-threading* y control de recursos: artificios pensados para esconder la **latencia**; resulta artificial en un entorno inherentemente *single-thread*.
- Nuevas **tecnologías**: grandes capacidades de memoria, “*hot standbys*”, la Web, etc.

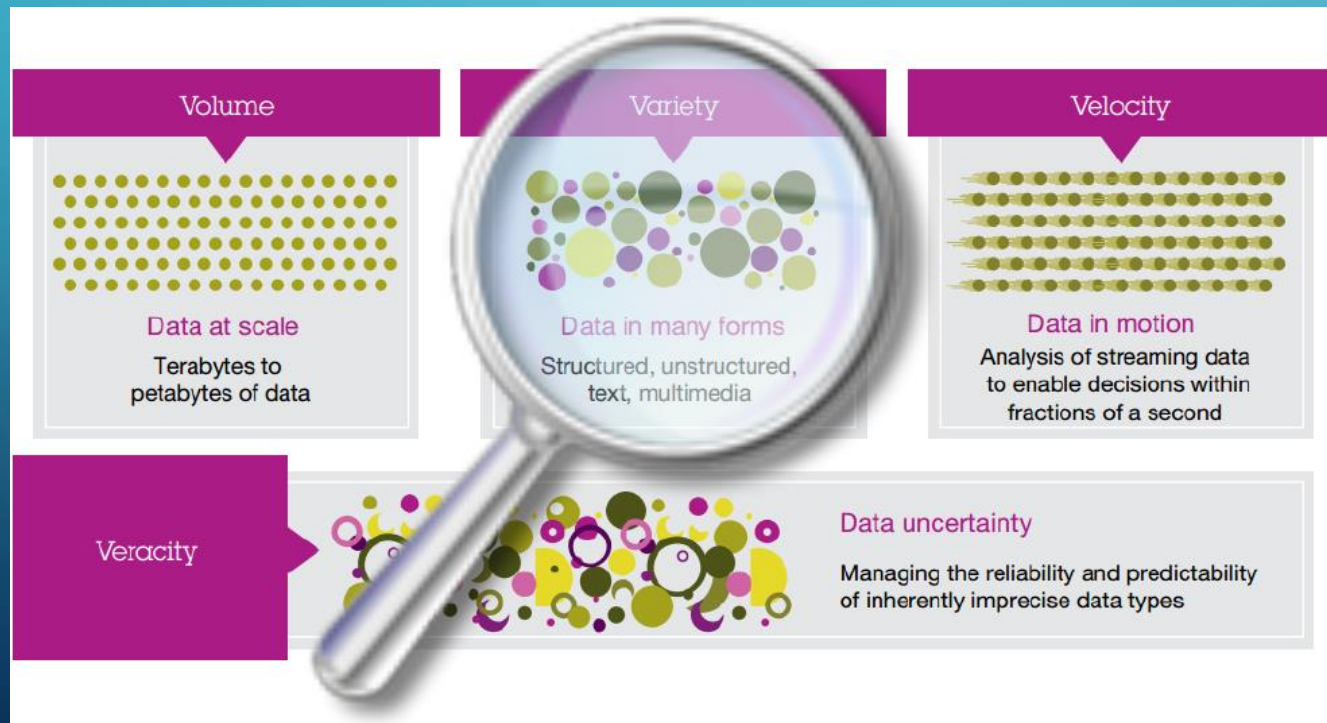
Velocidad: ¿Adiós a los DBMS?

Esto redundo en:

- La **caída** del modelo “*one size fits all*”: los diferentes problemas de manejo de datos se pueden resolver con arquitecturas de software **especializadas**.
- La falla de las implementaciones del **modelo relacional** para los mercados modernos.
- La necesidad de repensar tanto los **modelos de datos** como los **lenguajes de consulta**; cada aplicación especializada tiene la posibilidad de tomar las decisiones más convenientes.

Las 4 Vs del Big Data: *Variedad*

- Los **datos** son en general manifestaciones de **eventos**.
- El objetivo es **procesar** dichos eventos...

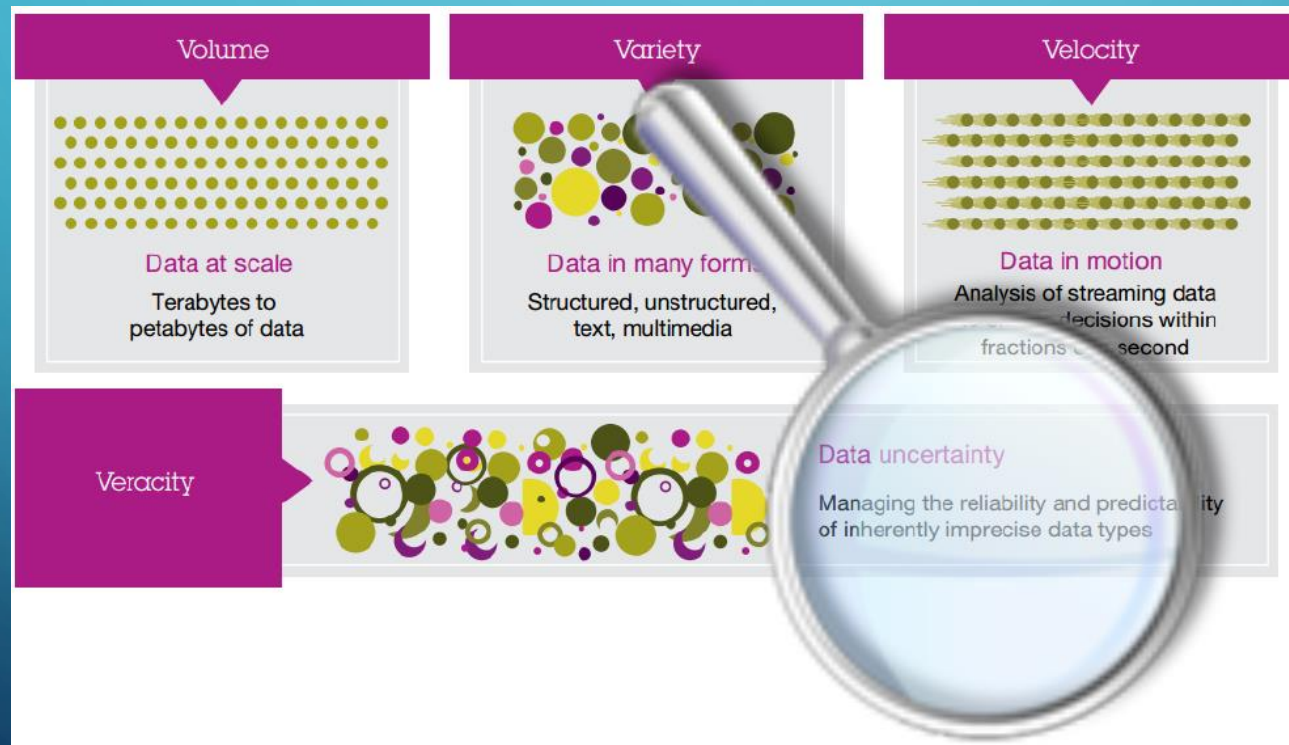


Variedad: *Eventos* (ej.)

- Los datos surgen de múltiples **fuentes**, y por lo tanto se encuentran representados en diferentes **formatos**.
- La **Web Semántica** se propuso como solución a este problema: el conocimiento se representa uniformemente:
 - RDF
 - OWL
 - SPARQL
 - Ontologías: lógicas de descripción, Datalog+/-, etc.
- Lamentablemente, el mundo no es estático, lo cual complica la aplicación de estas tecnologías...

Las 4 Vs del Big Data: *Veracidad*

La información proveniente de flujos generalmente es *incierta* por una o más razones.



Fuente: [DCM15]

Veracidad: *Soporte para incertidumbre*

- Muchas aplicaciones deben lidiar con fuentes de datos imprecisas o poco confiables:
 - Sensores ruidosos
 - Incompletitud
 - Inconsistencia
 - Información inherentemente incierta (clima, mercados financieros, etc.)
- La habilidad de **cuantificar** la incertidumbre asociada con la información es una parte esencial de la expresividad.
- Dos aspectos **ortogonales**:
 - Soporte para entradas con incertidumbre
 - Soporte para salidas con incertidumbre



20%

Structured Data



80%

Unstructured Data

PDFS

WORD DOCUMENTS

SPREADSHEETS

PRESENTATIONS

SOCIAL MEDIA POSTS

BOOKS