

# *Inteligencia de Negocios*

*Business Intelligence*

## *Big Data*

## **BIG DATA**

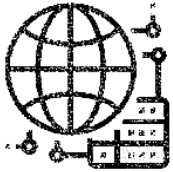
Los datos se expanden  
hasta llenar el espacio disponible



*Ley de los datos de Parkinson*

Big Data «es un término aplicado a conjuntos de datos que **superan la capacidad del software habitual** para ser capturados, gestionados y procesados en un tiempo razonable.

Los tamaños del “Big Data”  
se hallan constantemente en aumento»



Conectividad



Crecimiento  
exponencial  
de datos



Diferentes  
estructuras  
de Datos



Evolución  
de la  
Arquitectura  
de Datos





This is a highly detailed and colorful illustration representing the concepts of digital technology, innovation, and interconnected networks. The scene is set against a white background, featuring a variety of stylized elements:

- Buildings:** Several tall, colorful buildings in shades of red, orange, blue, and green are scattered across the landscape. Some buildings have unique features like a search bar, a clock, or a question mark.
- Icons and Symbols:** The illustration is filled with numerous icons, including gears, lightbulbs, speech bubbles, Wi-Fi symbols, SMS icons, and various charts and graphs. These symbols represent different aspects of technology and communication.
- Networks:** Dotted lines and arrows connect various points throughout the image, creating a sense of flow and interconnection. This network-like structure suggests a global or digital infrastructure.
- Landscape:** The bottom of the image shows a green, hilly landscape with small trees, a bridge, and a few tiny figures of people, grounding the abstract technological elements in a physical environment.
- Color Palette:** The color palette is bright and varied, using primary colors (red, blue, yellow) along with more muted tones (teal, grey) to create visual interest and distinguish between different components.

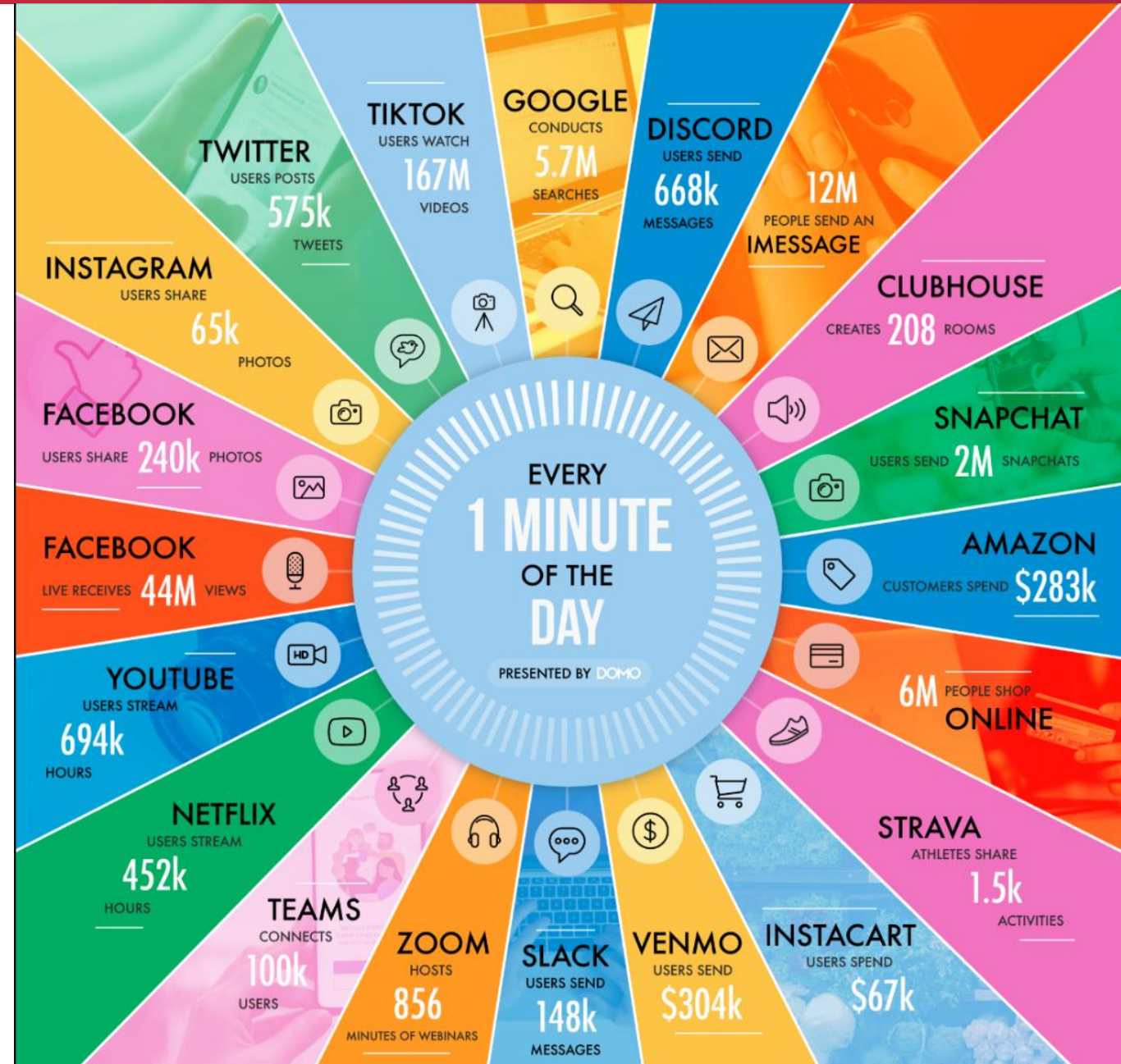
Overall, the illustration conveys a message of rapid technological advancement and the complexity of modern digital systems.

# Un minuto de Internet en 2021

Fuente: [Domo](https://www.domo.com/learn/infographic/data-never-sleeps-9) | [Visual Capitalist](https://www.domo.com/learn/infographic/data-never-sleeps-9) |

<https://www.domo.com/learn/infographic/data-never-sleeps-9>

Prof. Rita de la Torre





España **(+41%)**

~350 Gb/s



~700 Gb/s



Países Bajos **(+20%)**

~5 Tb/s



~7 Tb/s



Alemania **(+11%)**

~6 Tb/s



~7 Tb/s



## ESPECTADORES

**90 %** que buscan y leen  
...y con su "pasividad" generan datos

## PROSUMIDORES:

**8%** que consumen, opinan y conversan.  
...y con su actividad generan datos

## INFLUENCERS:

**2%(≈)** que influyen sobre el resto.  
...y también generan datos





#### ESTRUCTURADOS

- Tienen perfectamente definida su longitud, formato y tamaño.
- Se almacenan en formato tabla en Bases de Datos relacionales, archivos csv o en hojas de cálculo.



#### NO ESTRUCTURADOS

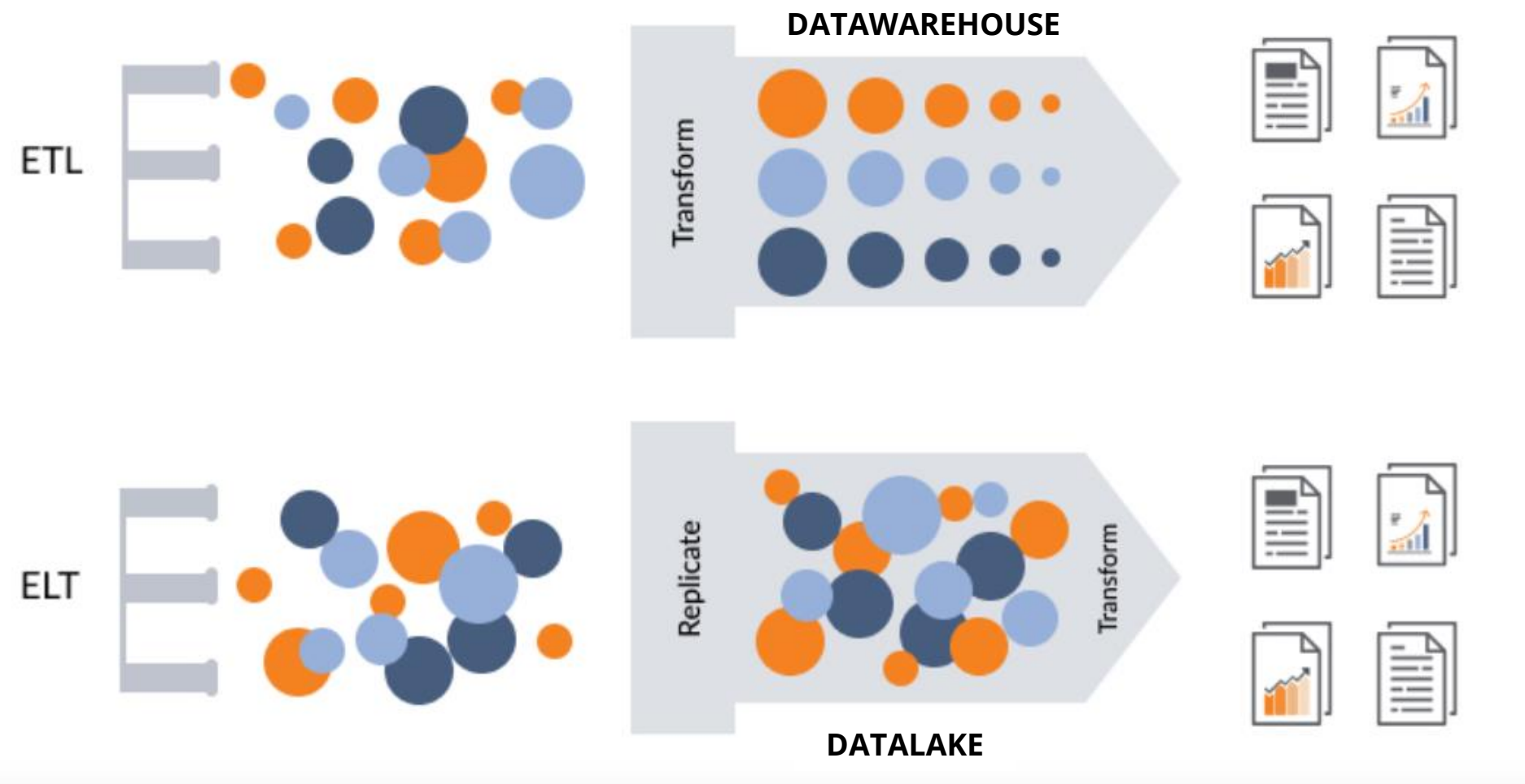
- Se caracterizan por no tener un formato específico
- Se almacenan en múltiples formatos como documentos PDF o Word, correos electrónicos, ficheros multimedia de imagen, audio o video.



#### SEMI ESTRUCTURADOS

No presentan una estructura perfectamente definida pero si presentan una organización definida en sus metadatos donde describen los objetos y sus relaciones, y que en algunos casos están aceptados por convención, como por ejemplo los formatos HTML, XML o JSON.

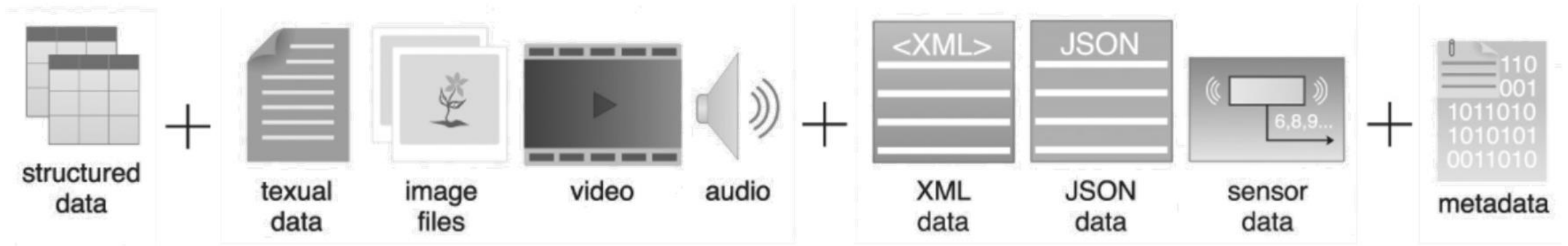
## ETL vs ELT



*Un **Data Lake** es un almacén para grandes cantidades de datos sin ningún tipo de tratamiento previo, es decir, sin procesar.*

Fuente: N-iX

<b>DATALAKE</b>	característica	<b>DATAWAREHOUSE</b>
estructurada semiestructurada desestructurada datos binarios (imágenes/vídeos)	tipo de datos	estructurados y procesados
para lectura, post-procesamiento	procesamiento	para escritura, pre-procesamiento
Generar procesos de Big Data, (Análisis en tiempo real y Machine Learning).	propósito	análisis para la toma de Decisiones de Negocio
Científicos de Datos e ingenieros	usuarios	Analistas de Negocio
Almacenamiento y analítica de datos, <i>deeplearning</i> y análisis en tiempo real	tareas	Consultas a los datos históricos para la realización de diversos análisis



Los metadatos proporcionan información sobre las características y la estructura de un conjunto de datos.

Este tipo de datos es en su mayoría generado por máquinas y se puede agregar a los datos.

El seguimiento de metadatos es crucial para el procesamiento, almacenamiento y análisis de Big Data porque proporciona información sobre los datos y su procedencia durante el procesamiento.

Ejemplos de metadatos incluyen:

- Etiquetas XML que proporcionan el autor y la fecha de creación de un documento
- Atributos que proporcionan el tamaño de archivo y la resolución de una fotografía digital

Las soluciones de Big Data se basan en los metadatos, cuando se procesan datos semiestructurados y no estructurados.



## Datos estructurados

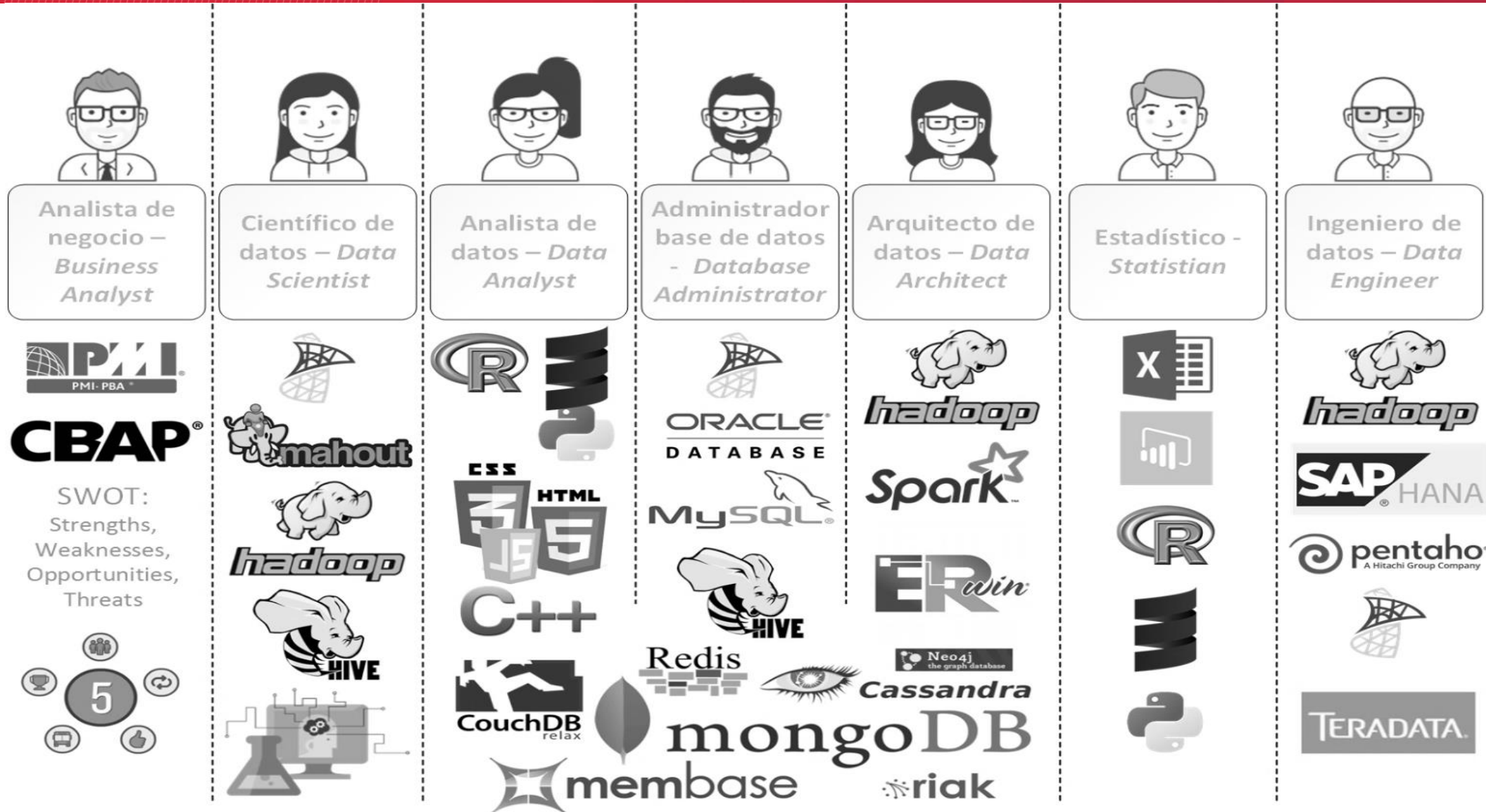
ID	Fecha	Usuario	IPv4 Address	Localización
1	2019-07-07	it08640	10.10.234.27	London
2	2019-07-07	rm05475	10.10.234.28	London
3	2019-07-07	gg04829	10.50.90.22	NY
4	2019-07-07	mx99922	10.10.234.30	London
5	2019-07-08	ss04472	10.10.234.31	London
6	2019-07-09	ui92212	10.50.90.24	NY
7	2019-07-10	mm93848	10.10.234.33	London

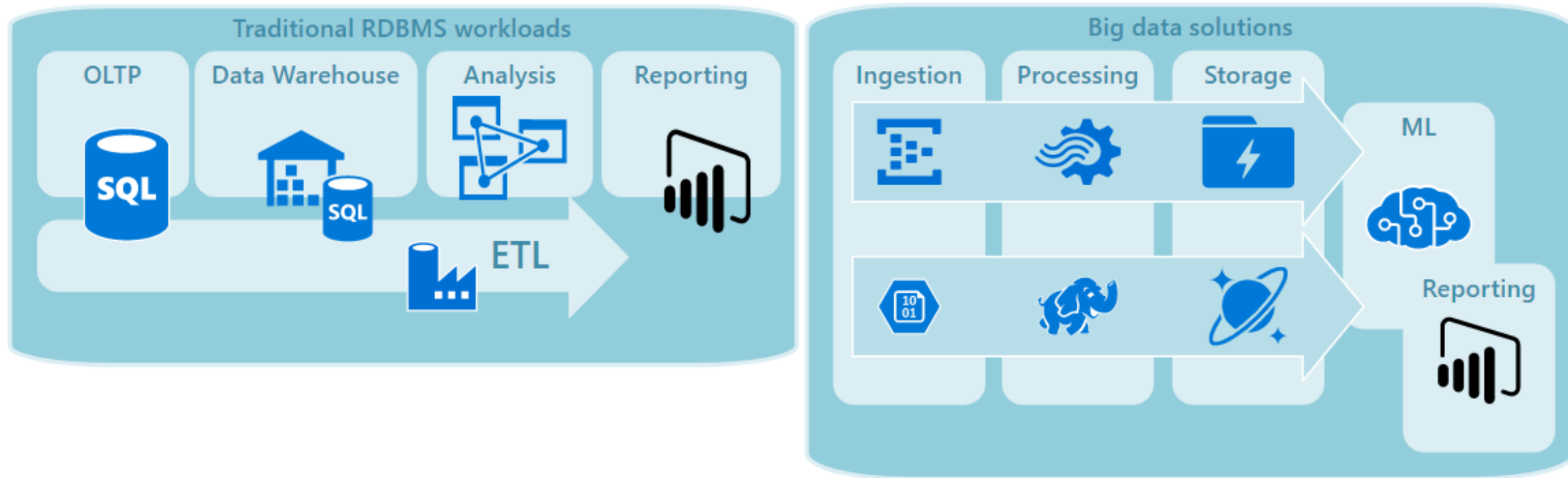
- ✓ Las filas y columnas estructuradas se adaptan a RDBMS
- ✓ Se estima que conforman un **20%** de los datos empresariales (Gartner)
- ✓ Más fácil de administrar con sistemas heredados

## Datos semiestructurados

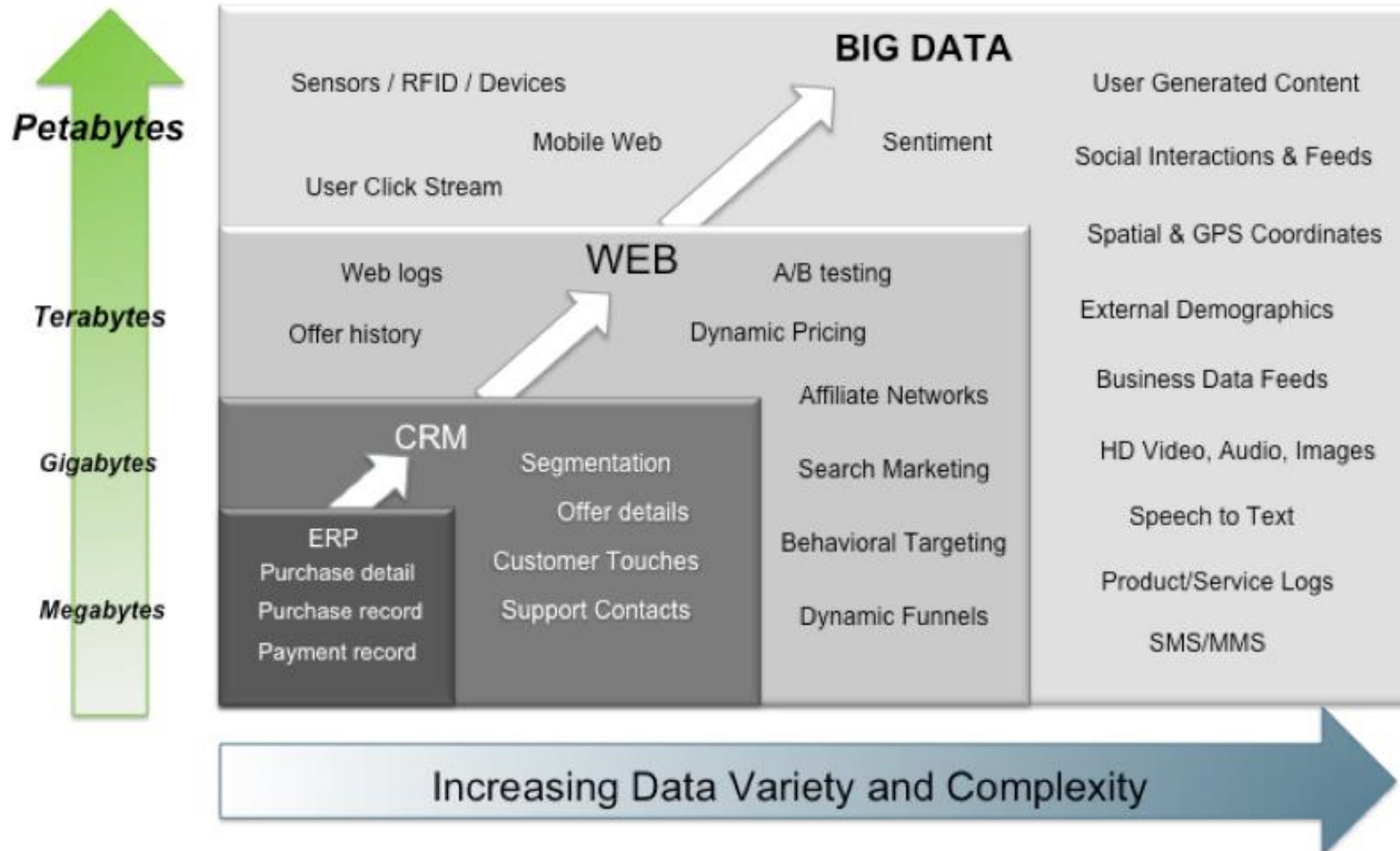
```
Jul 7 08:18:06 hpnod1 sshd[25428]: Failed password for root from  
::ffff:172.18.226.250 port 34087 ssh2  
Jul 7 08:18:06 hpnod1 sshd[25428]: Received disconnect from  
::ffff:172.18.226.250: 11: Bye Bye  
Jul 7 08:18:10 hpnod1 sshd[25430]: Failed password for root from  
::ffff:172.18.226.250 port 34229 ssh2  
Jul 7 08:18:10 hpnod1 sshd[25430]: Received disconnect from  
::ffff:172.18.226.250: 11: Bye Bye
```

- ✓ Los datos semiestructurados / no estructurados no se adaptan a las Bases de Datos Relacionales
- ✓ Se estima que son un **80%** de los datos empresariales (Gartner)
- ✓ Son difíciles de administrar con sistemas heredados

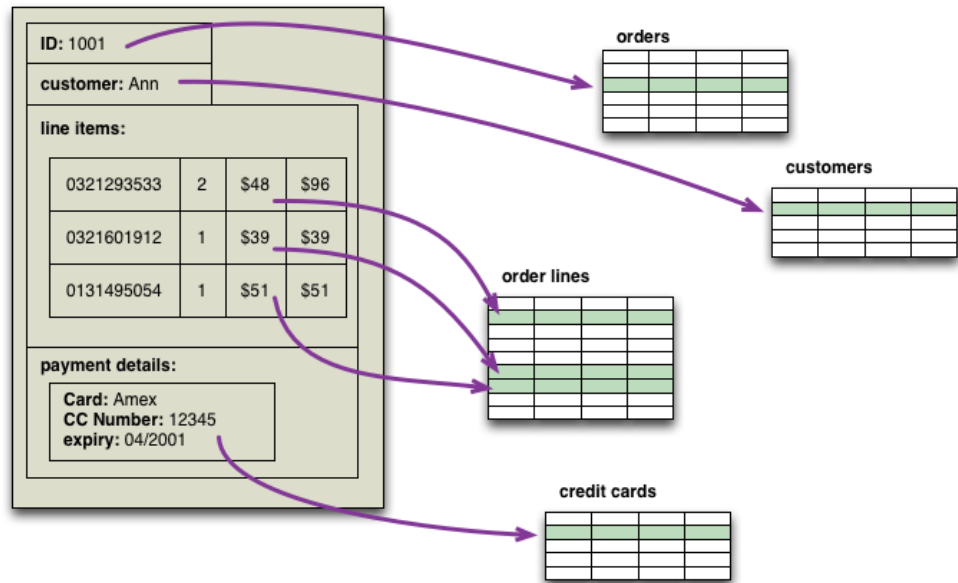




***Big Data*** = Transacciones + Datos Entorno + Interacciones + Observaciones



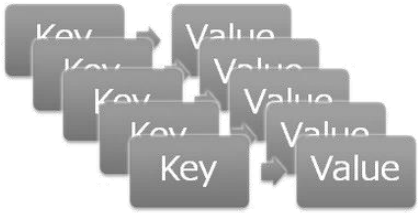




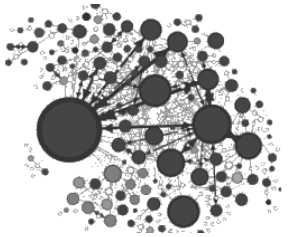
# SQL

Para operar con la información en las Bases de Datos **Relacionales** se utiliza **el lenguaje SQL**.

El lenguaje de consulta estructurado o **SQL** es un lenguaje declarativo estandar que permite manipular las bases de datos relacionales y especificar diversos tipos de operaciones en ellas.



Clave-Valor



Grafos



Familia de Columnas



Documentos

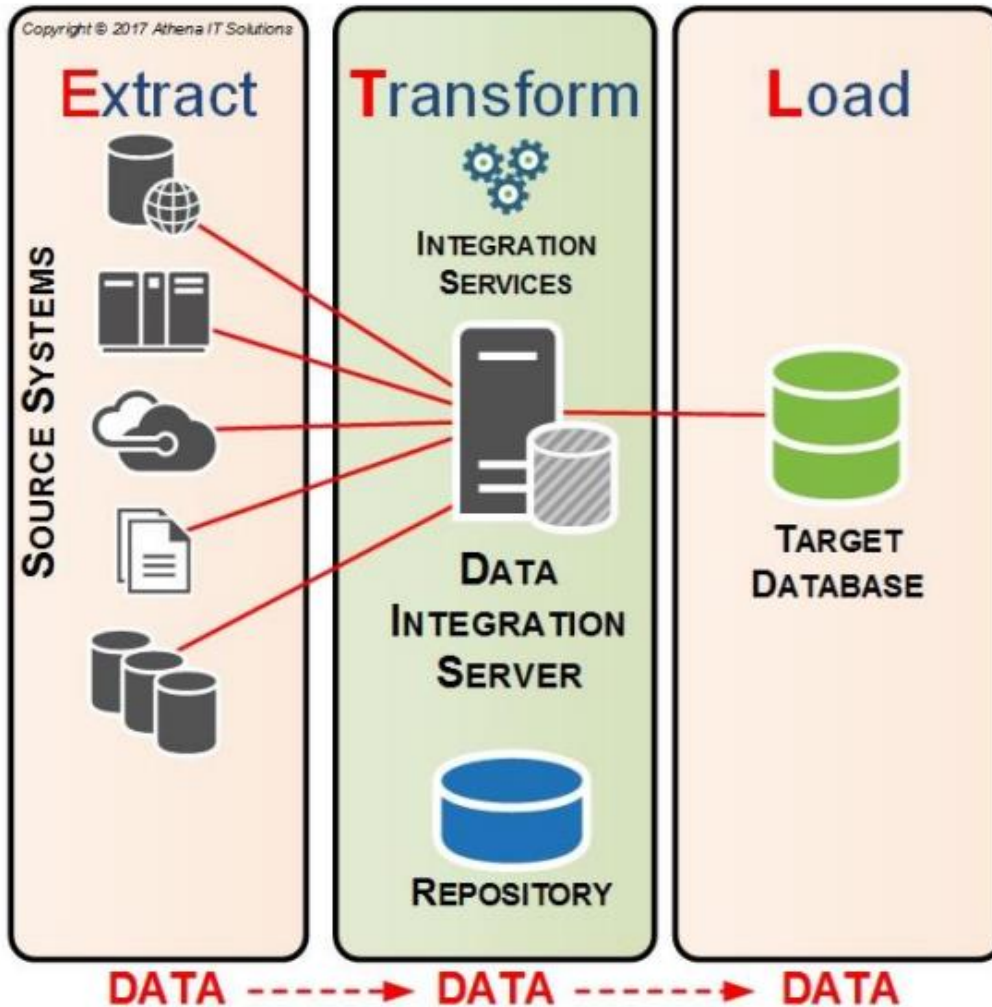
Bases de Datos de nueva generación que en su mayoría abordan **algunas** de las siguientes características:

- **no relacional**,
- distribuida,
- open-source, y
- horizontalmente escalable

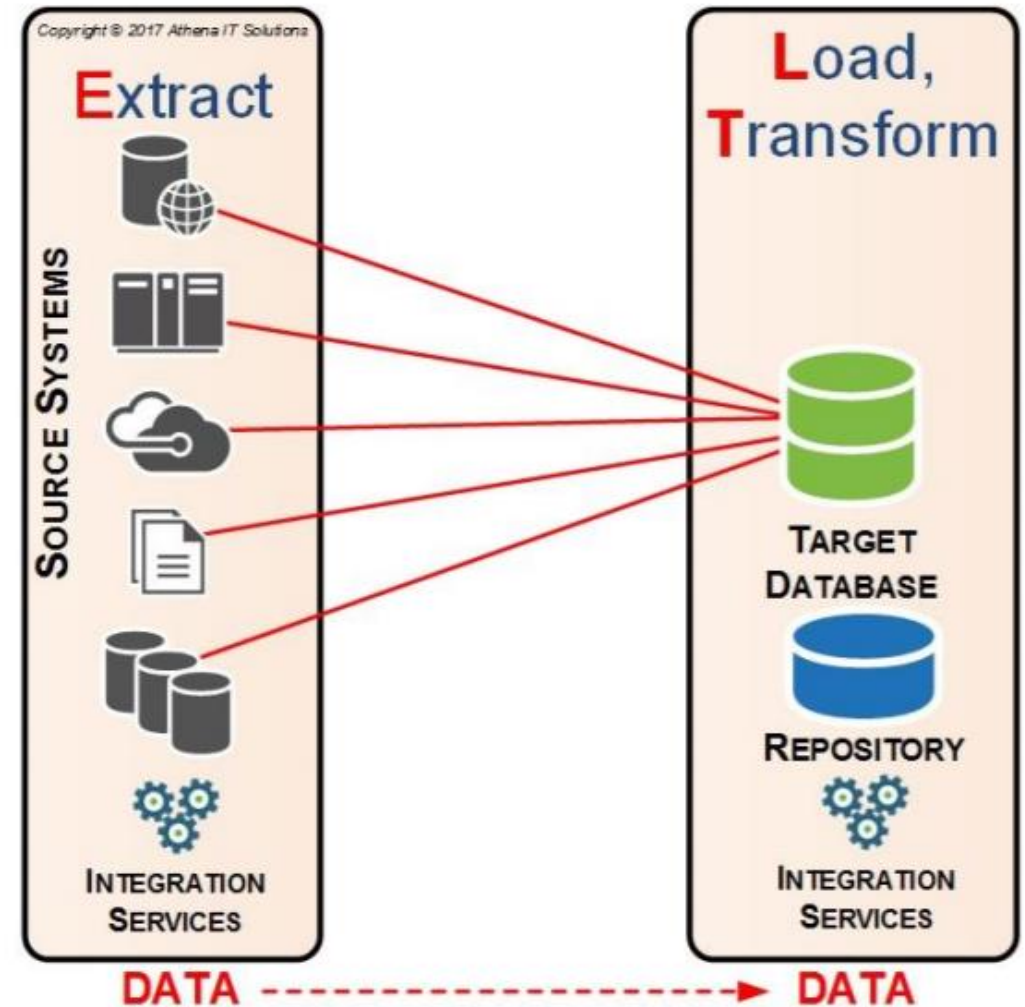
Utilizan otros lenguajes de consulta diferentes a SQL

**N**ot  
**O**nly SQL

## Extract, Transform & Load (ETL)



## Extract, Load & Transform (ELT)



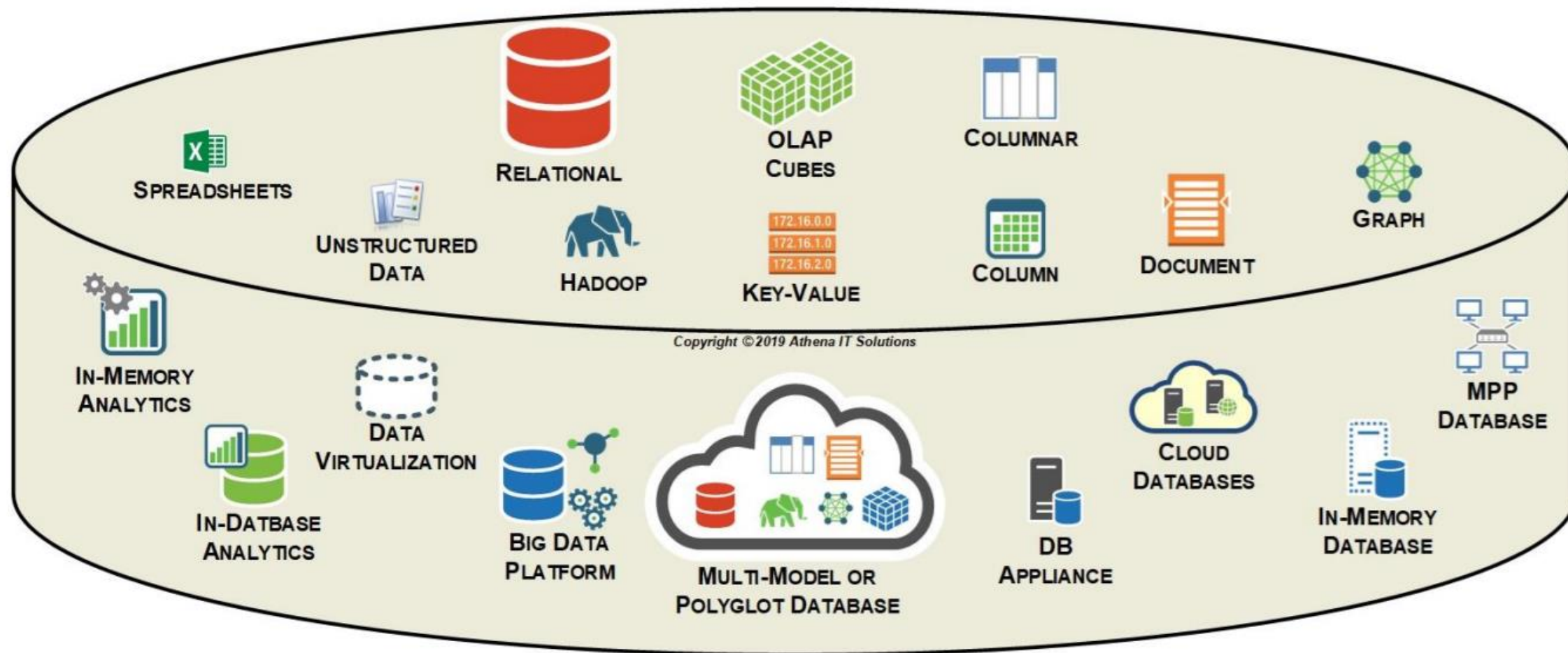


# Database Technologies

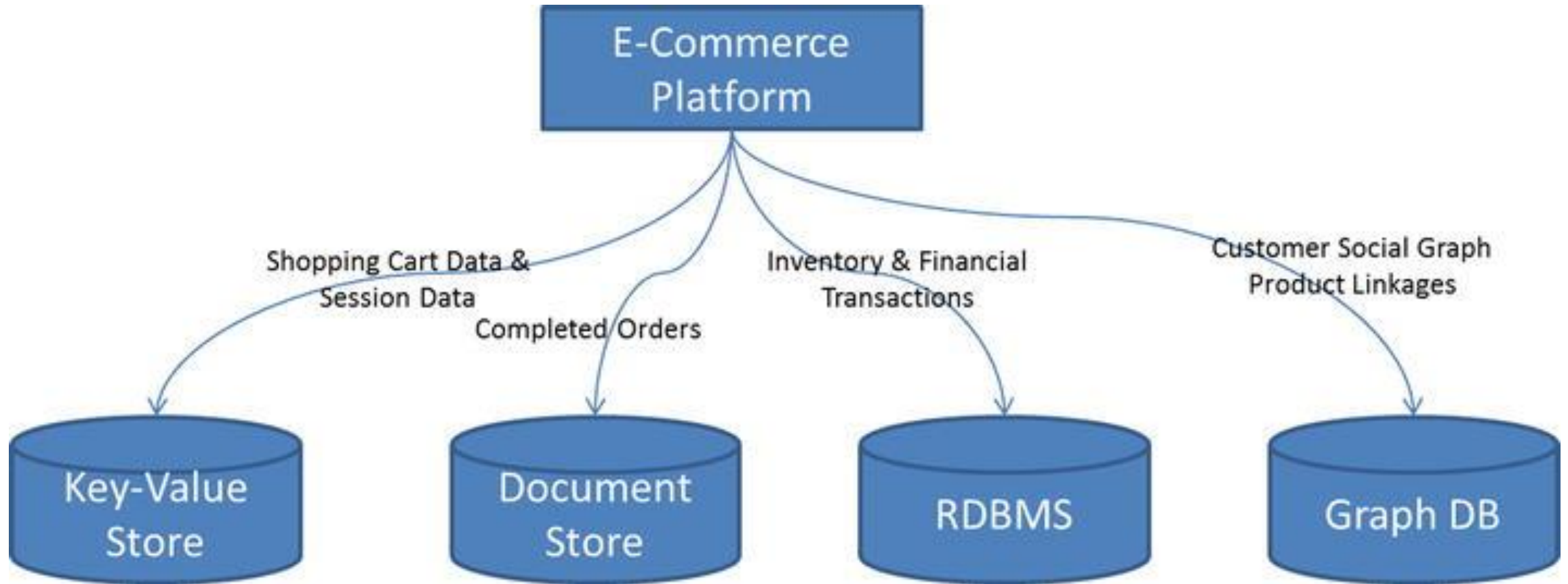


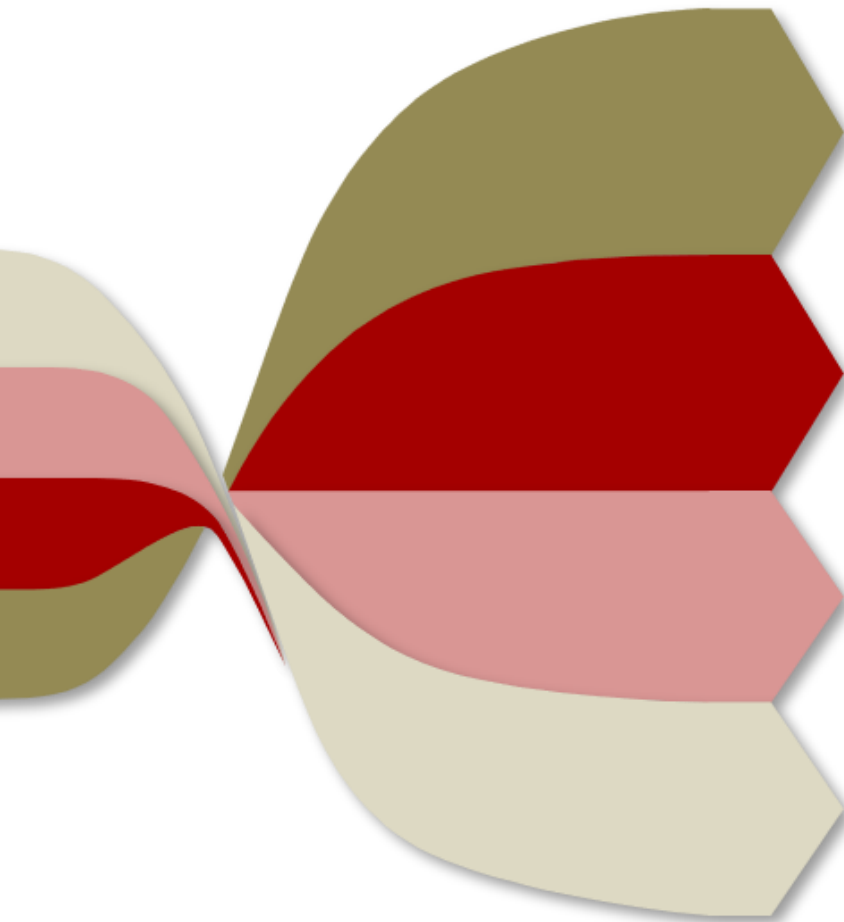
Google Cloud Platform Live

ORACLE









## **VOLUMEN**

Mayor cantidad de datos desde diferentes localizaciones

## **VELOCIDAD**

La digitalización ha aumentado la velocidad de la información

## **VARIEDAD**

Diversos tipos de datos provenientes de diferentes fuentes.

## **VERACIDAD**

Capacidad para obtener información a partir de conjuntos de datos alineados

## IMPLICACIONES

Las organizaciones deben implementar infraestructuras escalables para hacer frente al aumento de los volúmenes de datos

Los datos pueden ser recogidos en tiempo real.

Es necesario un modelo de datos flexible (NoSQL) para el aprovechamiento de la información.

Se ha de comprobar la validez en la información, ¿son los datos correctos y precisos para el uso previsto?

**VALOR** El BigData permite ayudar a la toma decisiones al momento, lo que se traduce en ventajas competitivas.



- Manejo de **diversas fuentes de información** estructurada y no estructurada
- Manipulación, transformación y visualización de **grandes volúmenes** de datos
- Uso de la información en **tiempo real**.
- Capacidad de **escuchar** más a nuestros **clientes**
- Desarrollo de **nuevos roles** en la empresa
- Elementos de juicio para las **decisiones** de la mejora operativa.



→ Aumentar los datos sobre los clientes y retener los más valiosos



→ Continuarmente mejorar la eficiencia operacional



→ Prevenir el fraude y gestionar el riesgo



→ Transformar y automatizar los procesos financieros

- **Optimización del marketing digital:**  
web analytics, attribution, golden path analysis
- **Exploración y descubrimiento de Datos:**  
Descubrir nuevos usos a los productos, nuevos mercados
- **Detección y prevención de fraude:**  
Integridad de sitios web
- **Análisis de Redes Sociales y relaciones:**  
Localización de influenciadores, posible *outsourcing*, predicción de mercados
- **Análisis de Retención de datos:**  
Datos de telecomunicaciones (registros de detalles de llamadas de telefonía, tráfico de internet y transacción de datos)



