

Webinars

Solex
2020

4. Computación Cloud

5. BI

6. Tableau Prep

7. Análisis de Sentimientos

8. Dpto. Dashboard

TEMAS

2020

The background features a dark blue field filled with numerous circles of varying sizes and shades of blue and teal. A faint, repeating watermark of the word 'SOLEX' is visible across the entire background. The text 'BIG DATA' is positioned in the lower right quadrant.

BIG DATA

BIG DATA

“Big Data trata de **entender las relaciones** que te unen a las personas que te interesan y el **conocimiento del potencial** de esa relación” **Joe Rospars**

Big data (Datos masivos) es un término que describe el **gran volumen de datos** – estructurados y no estructurados – que inundan una empresa todos los días. Pero no es la cantidad de datos lo importante. Lo que importa es lo que **las organizaciones hacen con los datos**.

REVOLUCIÓN DE LOS DATOS.

Sociedad industrial a una sociedad del conocimiento

Se están **generando** y **registrando** cantidades descomunales de información, tenemos datos del **comportamiento** humano que no habían existido hasta ahora, nuestros actos cotidianos están siendo observados, almacenados y estudiados como nunca antes había ocurrido.

REVOLUCIÓN DE LOS DATOS.

- El **gran volumen** de datos que se generan de forma **continuo** y los datos de los que ya se dispone.
- la **velocidad de procesamiento** que permite ya a analizar esa información.

PROCESOS SOBRE LA INFORMACIÓN (5 V)

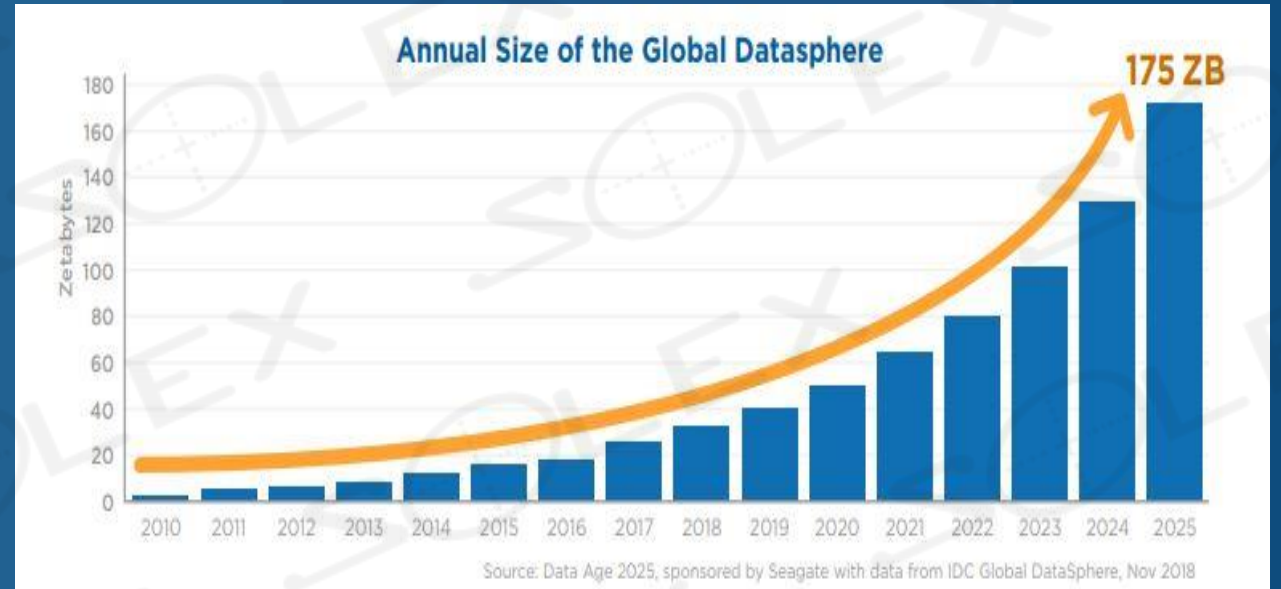
VOLUMEN

El volumen de datos llegará a **175 zettabytes** en **2025**, según un informe de la consultora IDC (*International Data Corporation*), lo que significa el equivalente a 175 veces la información generada en 2011.

2018 - **33 Zettabytes** aprox.

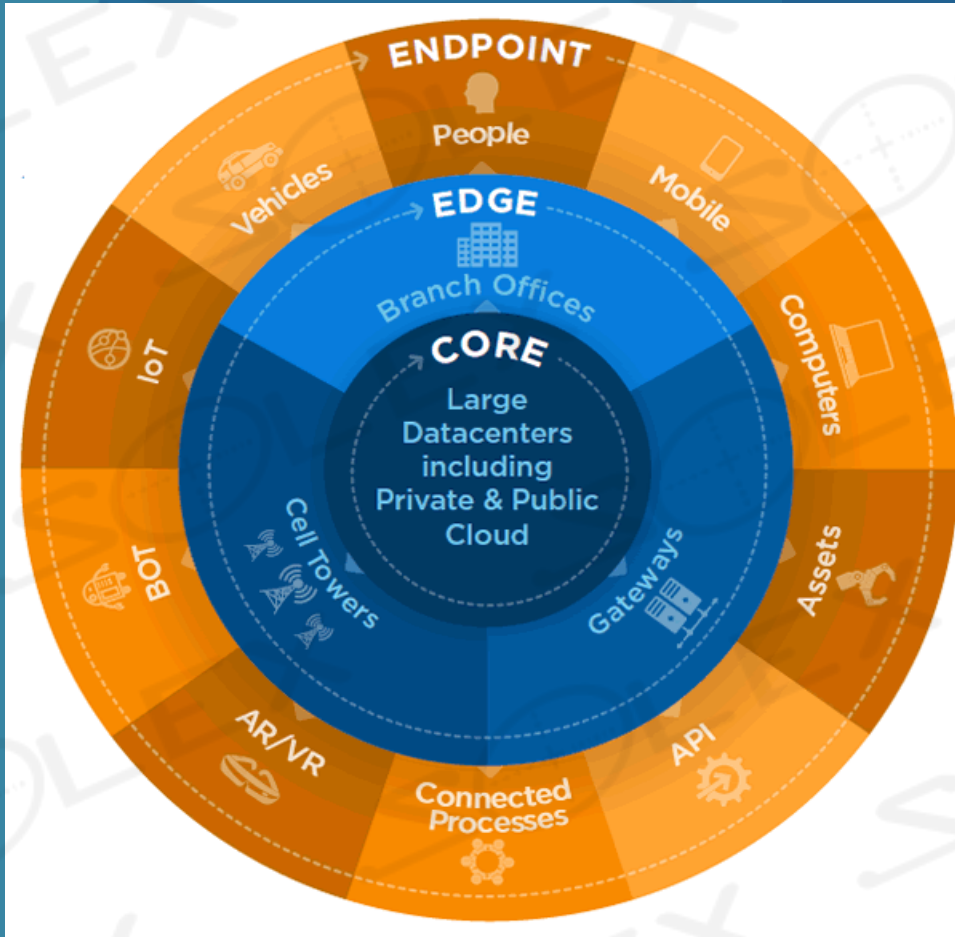
2020 - **44 Zettabytes** aprox.

- 1 zettabyte equivale a 1 trillon gigabytes - 10^{21} bytes



Fuente: The Digitization of the World From Edge to Core

VOLUMEN



Fuente: IDC's Data Age 2025 study, sponsored by Seagate

Core (núcleo): centros de datos informáticos designados en la empresa y proveedores de la nube.

Edge (Borde): se refiere a servidores y dispositivos reforzados para empresas que no están en los centros de datos centrales.

EndPoint: los dispositivos en el borde de la red, incluidos PC, teléfonos, dispositivos industriales sensores, automóviles conectados y wearables.

VOLUMEN (por minuto)

Tecnología e internet

- Se venden **2.833** teléfonos celulares
- Se envían **159.362.760** e-mails
- Skype soporta **176.220** llamadas
- Se envían **12.986.111** mensajes de texto
- Se producen **3.877.140** búsquedas en Google

Retail

- Amazon envía **1.111** paquetes y vende **\$332.876** dólares
- Se generan **73.249** transacciones por internet
- Los compradores dejan **\$7.610.350** dólares abandonados en sus carritos de compras



Media, entretenimiento y redes

- Se reproducen **97.222** horas de video en Netflix
- Los usuarios de Youtube miran **4.333.560** videos
- Se suben **400** horas de video a Youtube
- Se escriben **473.400** tweets en Twitter
- Se publican **49.380** fotos en Instagram
- Spotify reproduce más de **750.000** canciones

Es el **ritmo** en que los datos de entrada **fluyen** desde las diversas fuente.



VELOCIDAD

Para los procesos en los que el tiempo resulta fundamental, tales como la detección de fraude en una transacción bancaria o la monitorización de un evento en redes sociales, estos tipos de datos deben estudiarse en tiempo real para que resulten útiles para el negocio y se consigan conclusiones efectivas.

Lo rápido que hoy en día se puede distribuir una foto que tomamos a millones de personas por todo el mundo a través de las redes sociales o lo rápido que google nos devuelve información del otro lado del planeta a partir de nuestras búsquedas.

VARIEDAD

Se refiere a las **formas, tipos y fuentes** en las que se registran los datos.

Estructurados:

patrones específicos para ser almacenada, nos referimos a la información que se suele encontrar en la mayoría de bases de datos.

se adaptan a un modelo de datos predefinido. Se organizan de forma tabular con relaciones predecibles y constantes entre filas y columnas.



VARIEDAD

Se refiere a las **formas, tipos y fuentes** en las que se registran los datos.

No Estructurados:

Son datos binarios que no tienen estructura interna identificable.

- Correos electrónicos.
- Archivos de procesador de texto.
- Archivos PDF.
- Hojas de cálculo.
- Imágenes digitales.
- Vídeo.
- Audio.
- Publicaciones en medios sociales.

Datos no estructurados



El **80 %** de la información relevante para un negocio se origina en forma no estructurada, principalmente en formato texto.

VARIEDAD

Se refiere a las **formas, tipos y fuentes** en las que se registran los datos.

Semi Estructurados:

Son datos ordenados que no se adecúan a los modelos de datos tabulares. .

```
- <experiments version="1.2" revision="100915" total="58" total-samples="4011" total-assays="3847">
- <experiment>
  <release date>2007-11-22</release date>
  <species>Mus musculus</species>
  - <miamescores>
    <reporter sequence score>1</reporter sequence score>
    <factor values score>1</factor values score>
    <measured bioassay data score>0</measured bioassay data score>
    <protocols score>0</protocols score>
    <derived bioassay data score>1</derived bioassay data score>
    <overall score>3</overall score>
  </miamescores>
  <assays>18</assays>
  <samples>18</samples>
  <raw data files>0</raw data files>
  <file data files>18</file data files>
  - <sample attribute>
    <category>CellType</category>
    <value>primary chondrocyte</value>
    <value>primary dermal fibroblast</value>
    <value>primary osteoblast</value>
  </sample attribute>
  - <sample attribute>
    <category>Organism</category>
    <value>Mus musculus</value>
  </sample attribute>
  - <experimental factor>
    <name>CellType</name>
    <value>primary chondrocyte</value>
    <value>primary dermal fibroblast</value>
  </experimental factor>
</experiment>
</experiments>
```

Tienen metadatos que los describen. Por ejemplo, los datos asociados a un mensaje de texto que describen cuándo se envió, cuándo fue recibido, etc.

En general estos metadatos vienen en formato JSON o XML y muchas soluciones de Big Data se adaptan para leerlos y procesarlos.

VERACIDAD



Incertidumbre de los datos, es decir, al grado de **fiabilidad** de la información recibida. Sesgo, el ruido y la alteración de datos.

medida directa de su calidad. (**Disponibilidad, Usabilidad, Confiabilidad, Pertinencia, Calidad de presentación**)

limpieza de los datos.

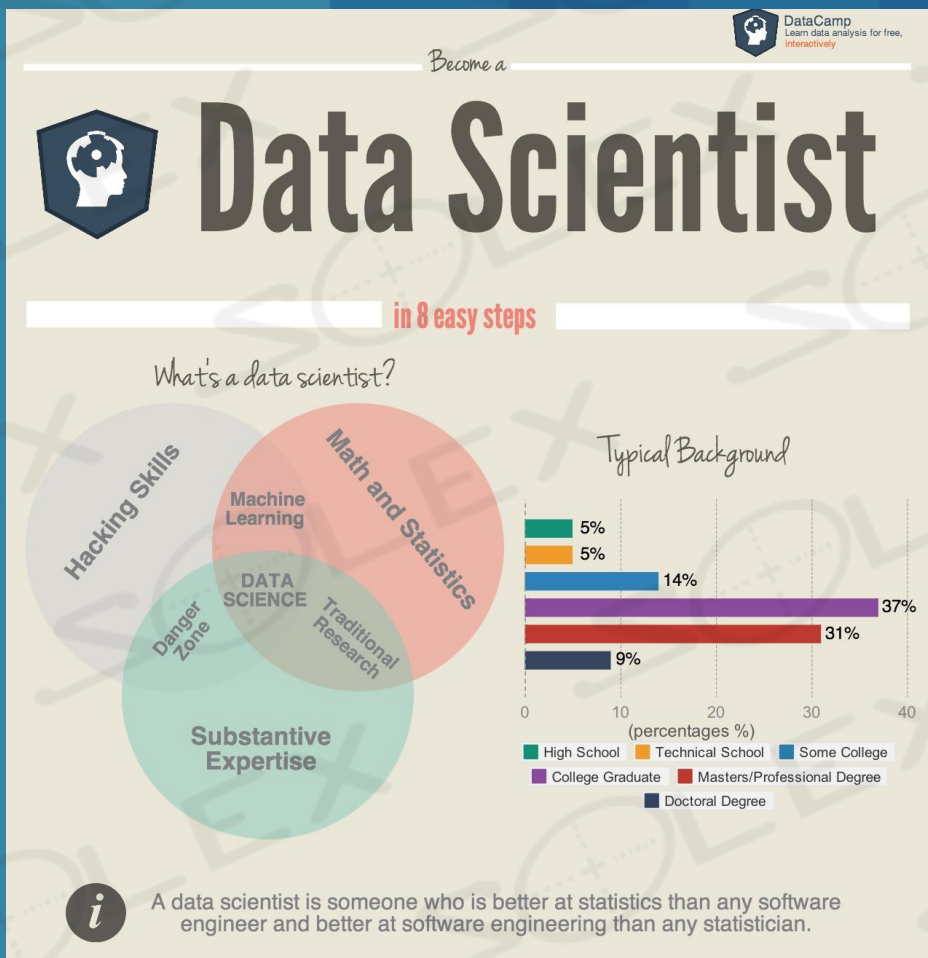
Validez

¿son los datos **correctos** y **precisos** para el uso previsto?

Volatilidad

Se refiere al tiempo durante el cual los datos recogidos son válidos y el **tiempo** que deben permanecer almacenados.

VALOR



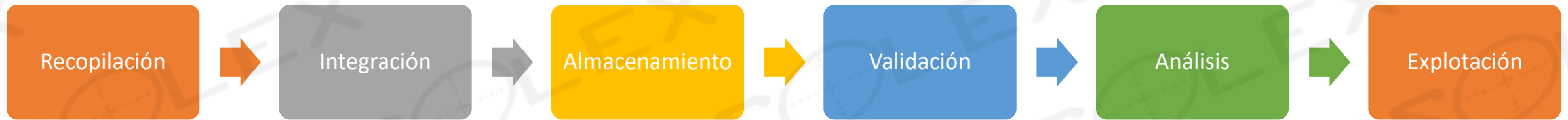
El valor se obtiene de datos que se transforman en **información**; esta a su vez se convierte en **conocimiento**, y este en acción o en **decisión**.

tecnologías de procesamiento de lenguaje natural **(PLN)**

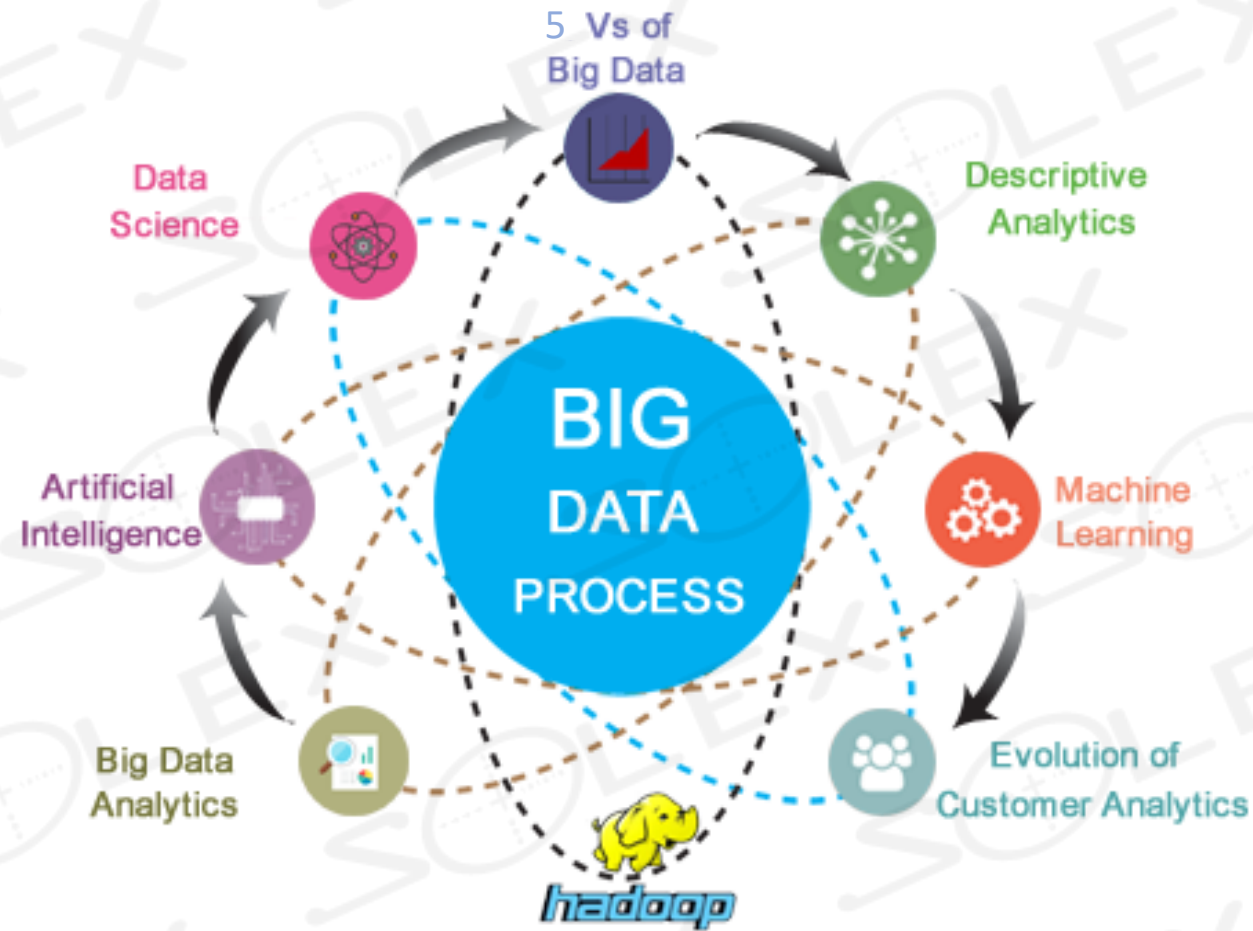
- Detector automático de análisis del sentimiento
- Clasificador automático de documentos
- Detector de topics
- Detector de similitudes en documentos
- Detector de Named-entity recognition (NER)
- Detector de género y edad en redes sociales

Fuente: <https://machinelearningmastery.com/become-data-scientist/>

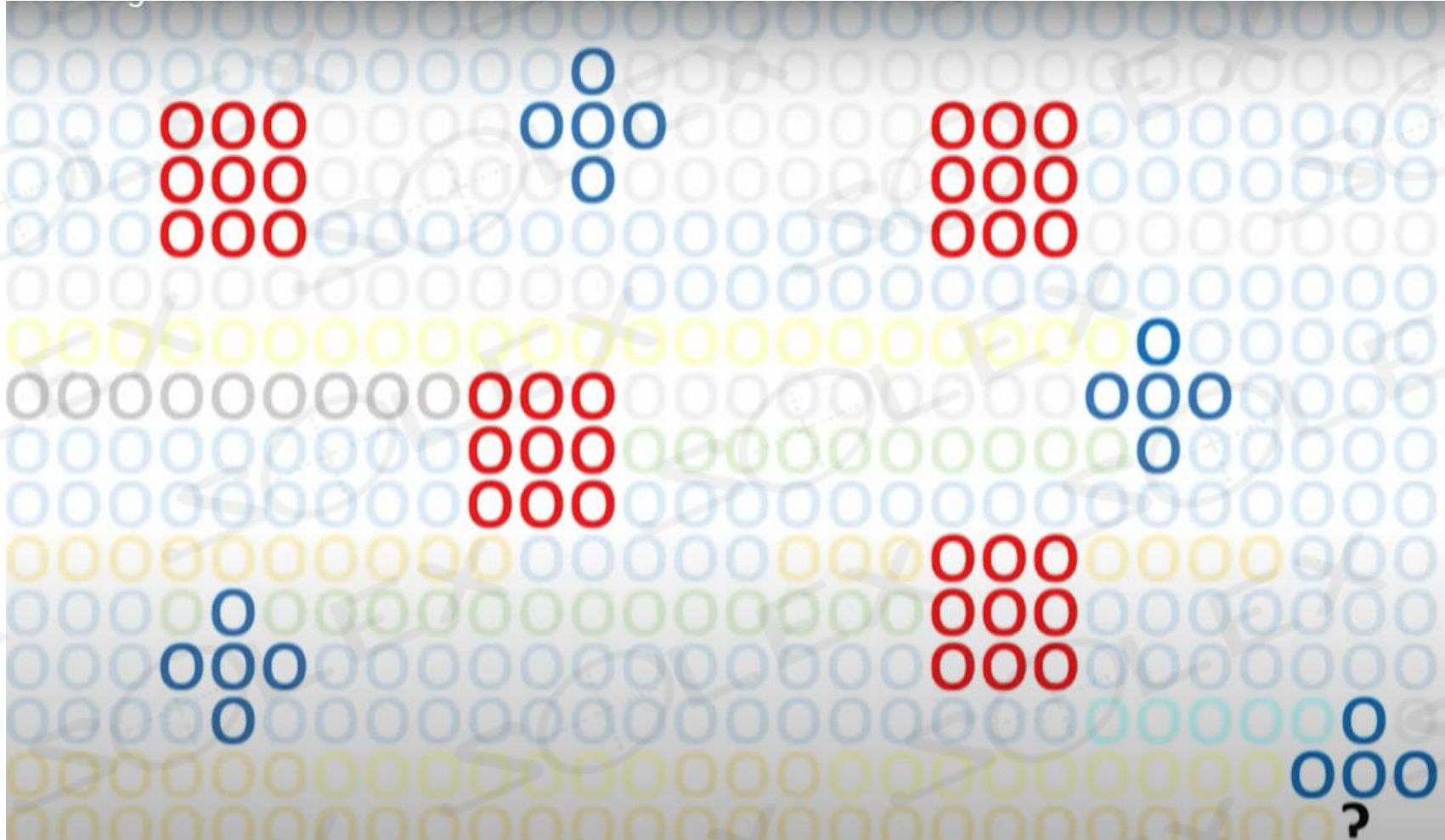
Procesos



Disciplinas de interpretación



Disciplinas de interpretación



Minería De Datos



Proceso analítico diseñado para explorar datos (usualmente grandes cantidades) en busca de patrones, tendencias y comportamientos consistentes entre variables.

Modelado descriptivo, Modelado predictivo, Modelado prescriptivo

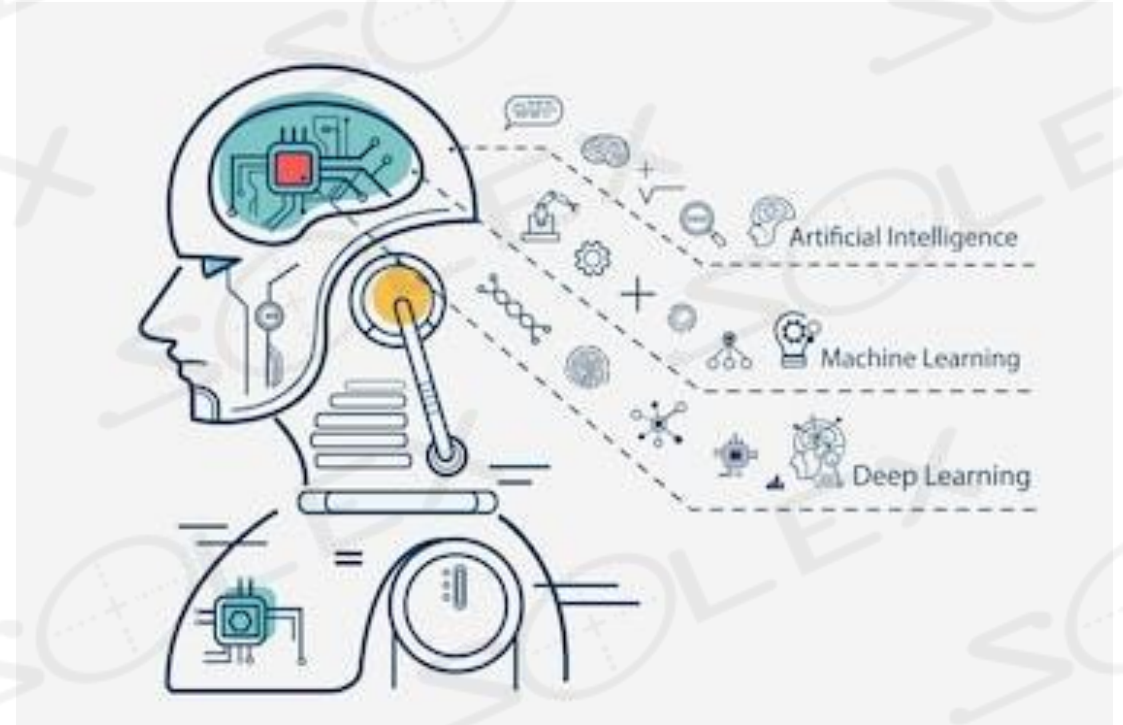
Inteligencia Artificial



Fuente: <https://signalsiot.com/asi-es-como-el-big-data-y-la-inteligencia-artificial-pueden-apoyar-a-las-pymes/>

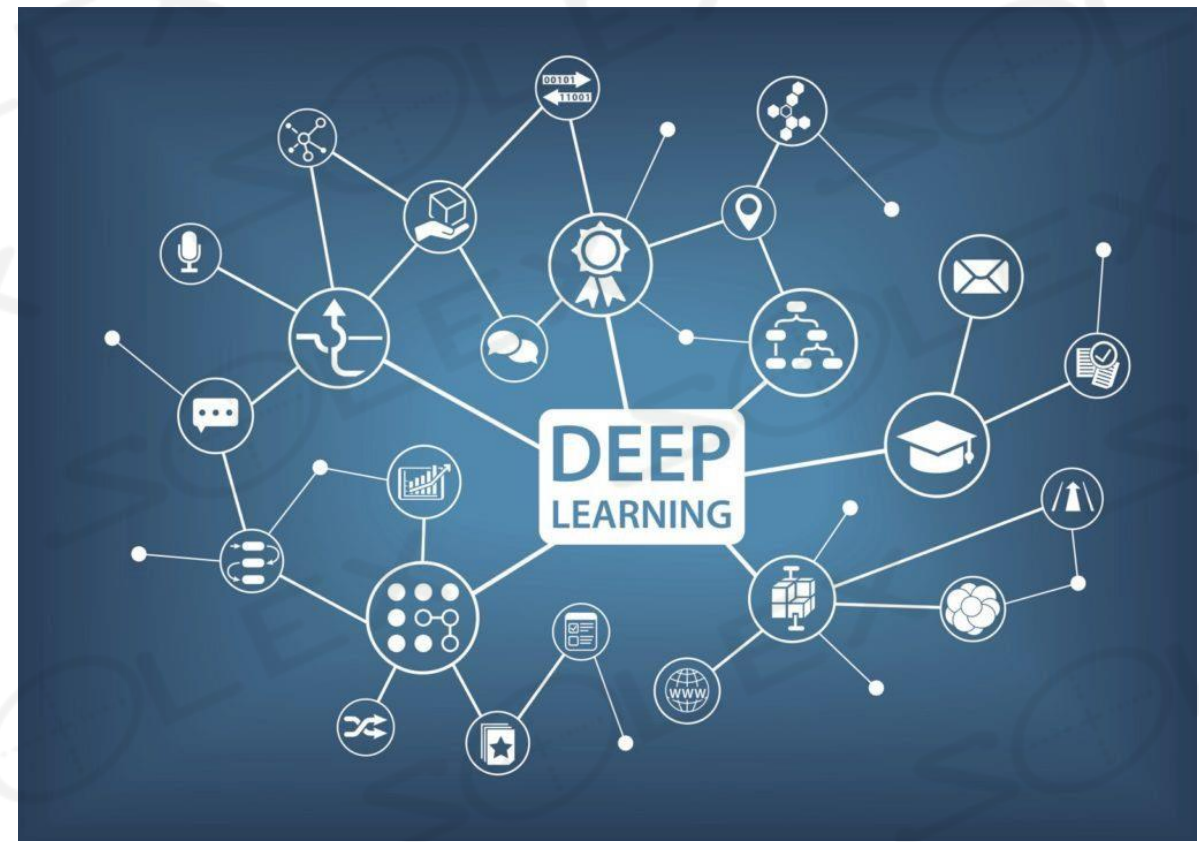
Hace posible que las máquinas aprendan de la experiencia, se ajusten a nuevos escenarios y realicen tareas como seres humanos. La mayoría de los ejemplos de inteligencia artificial sobre los que oye hablar hoy día – desde computadoras que juegan ajedrez hasta automóviles de conducción autónoma

Aprendizaje Autónomo (Machine Learning)



Automatizar, mediante distintos algoritmos, la identificación de patrones o tendencias que se “esconden” en los datos.

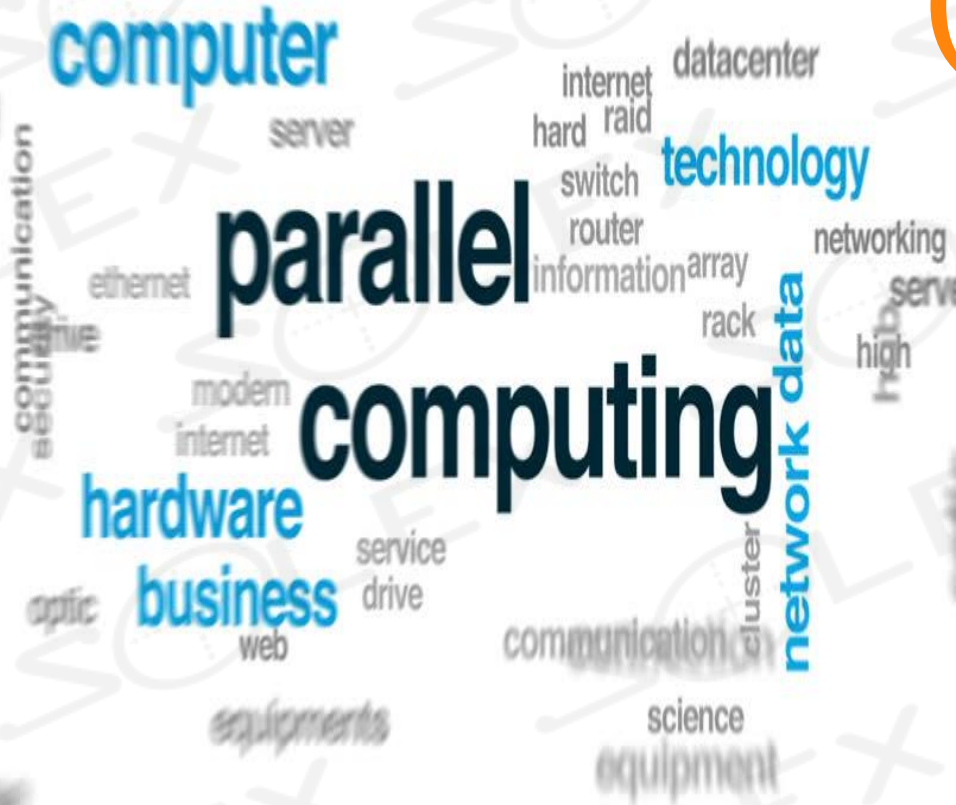
Aprendizaje Profundo (Deep Learning)



Fuente: <https://www.profesionalreview.com/2019/08/24/deep-learning-que-es-y-como-se-relaciona-con-el-machine-learning/>

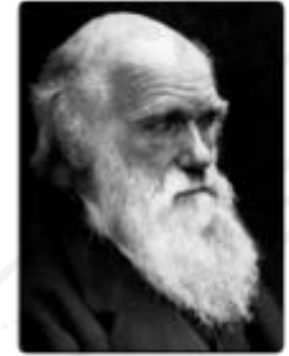
Proceso de Machine Learning usando una red neuronal artificial que se compone de un número de niveles jerárquicos.

Computación en Paralelo



La computación paralela es una técnica de programación en la que muchas instrucciones se ejecutan simultáneamente.

Computación Evolutiva



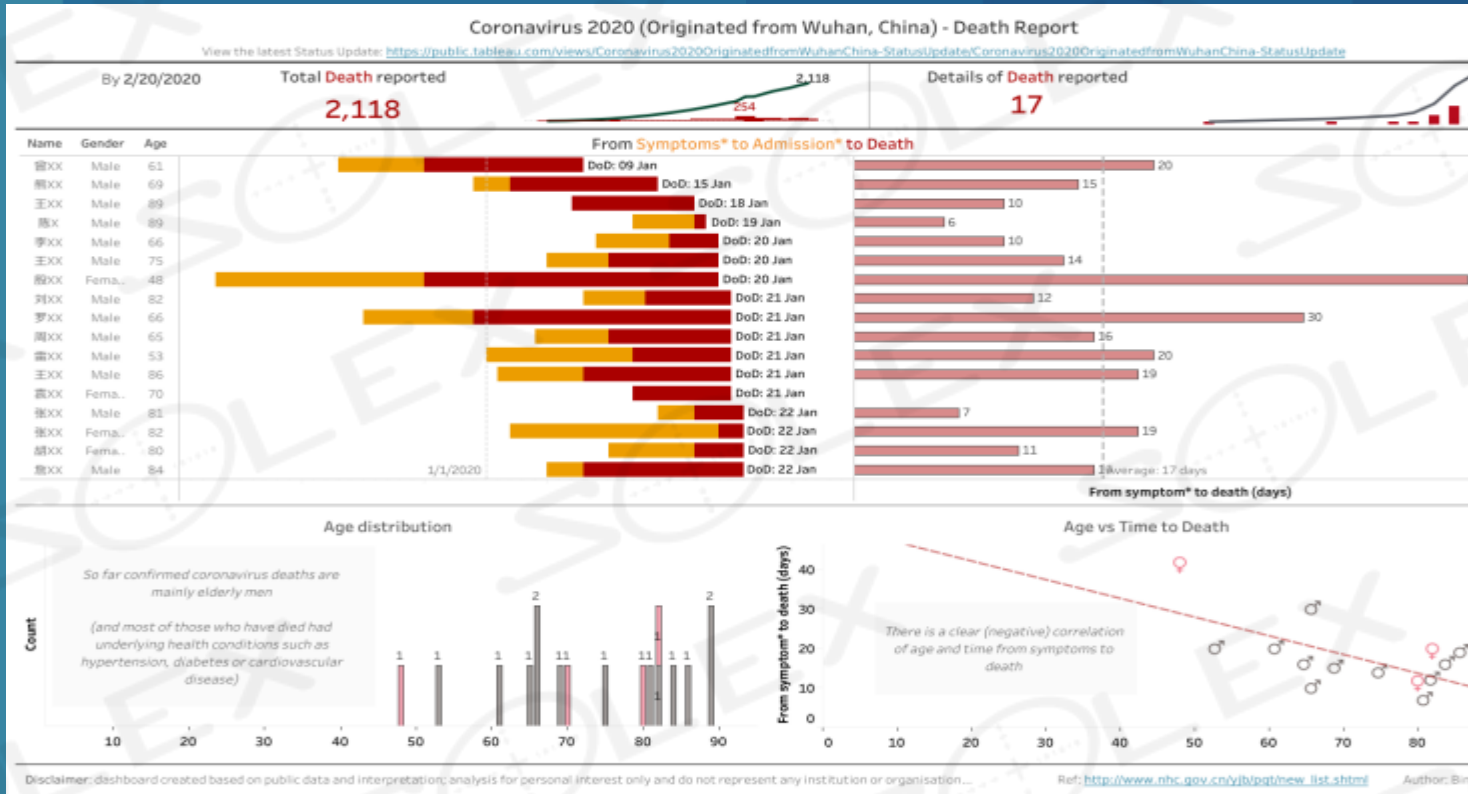
Darwin, Medel y Lamark.

Naturaleza	Algoritmos Evolutivos
Individuo	Solución al problema
Población	Conjunto de soluciones
Adecuación(fitness)	Calidad de la solución
Cromosoma	Representación o codificación de una solución
Gen	Parte o componente de la representación de una solución
Mutación y recombinación	Operadores de búsqueda
Selección natural	Preservación o re-utilización de buenas soluciones (o sus componentes)

interpreta la naturaleza como una inmensa máquina de resolver problemas y trata de encontrar el origen de dicha potencialidad para utilizarla en nuestros programas.

VISUALIZACIÓN

Modo en el que los datos son **presentados**. de manera que sean **legibles** y **accesibles**, para encontrar patrones y claves ocultas en el tema a investigar.



“Comenzamos con un caso de uso corporativo pequeño, que luego comenzó a expandirse rápidamente. Mientras todos quieren hablar del análisis de big data, Tableau lo que quiere es simplificarlo”. — **ASHISH BRAGANZA, DIRECTOR DE INTELIGENCIA DE NEGOCIOS INTERNACIONAL, LENOVO**

Aplicaciones

- El cambio **climático**
- **Políticas** Publicas.
- Curar **enfermedades**
- Yacimientos petrolíferos
- Optimizar inversiones
- Investigaciones **genéticas** –
Nanotecnología – implantaciones en el cuerpo.
- **Northpointe** Inc. – Cuando un preso puede reincidir en Estados Unidos

Domótica e IoT

- Ahorro de Energía
- Compras
- Necesidades
- Seguridad

Smartcities

- Calidad del aire
- Centros de salud
- Control de tráfico
- Accidentes naturales
- Predicción Meteorológica
- Seguridad Ciudadana

The background features a pattern of overlapping circles in various shades of blue and teal. A large, semi-transparent teal circle is prominent at the top center. The word 'SOLEX' is repeated in a light, semi-transparent font across the background, often appearing inside the circles. The overall aesthetic is clean and modern.

Gracias