

Introducción a la Ciencia de Datos

Preng Biba



Agenda

1. Introducción al Curso.
2. Antes de Iniciar, un poco de Historia.
3. Hablemos de Inteligencia Artificial.
4. Evolución de Data Science.
5. Algo de Terminología.
6. ¿Que necesita saber un científico de Datos?
7. Aplicaciones y Nuevas Áreas.



- Ing. Preng Biba
 - Ingeniero en Mecatrónica,
 - Maestría en Electrónica Industrial,
 - Maestría en Investigación de Operaciones.

- Ing. Diego Chuy
 - Ingeniero en Mecatrónica,
 - Maestría en Electrónica Industrial,

Un poco de Historia:

- Inicia en 1760 hasta 1840.
- Primera computadora 1936.
- Primer transistor 1947 (nacimiento de la electrónica).
- Primera generación de computadoras (de 1945 a 1955).

Un poco de Historia:

- Primer Circuito integrado 1960.
- Sistemas operativos de usuario 1980.
- Auge del internet 1990.
- Facebook aparece en el 2004.

Un poco de Historia:

- Aparece Hadoop 2006.
- Plataformas de desarrollo para Data Science 2008.

Hablemos de
Inteligencia Artificial.

Inteligencia Artificial

“Dotar a un sistema con la capacidad de aprender, razonar y resolver problemas basandose en el comportamiento de los seres vivos.”

“Diseñar sistemas que resuelven tareas de forma inteligente.”

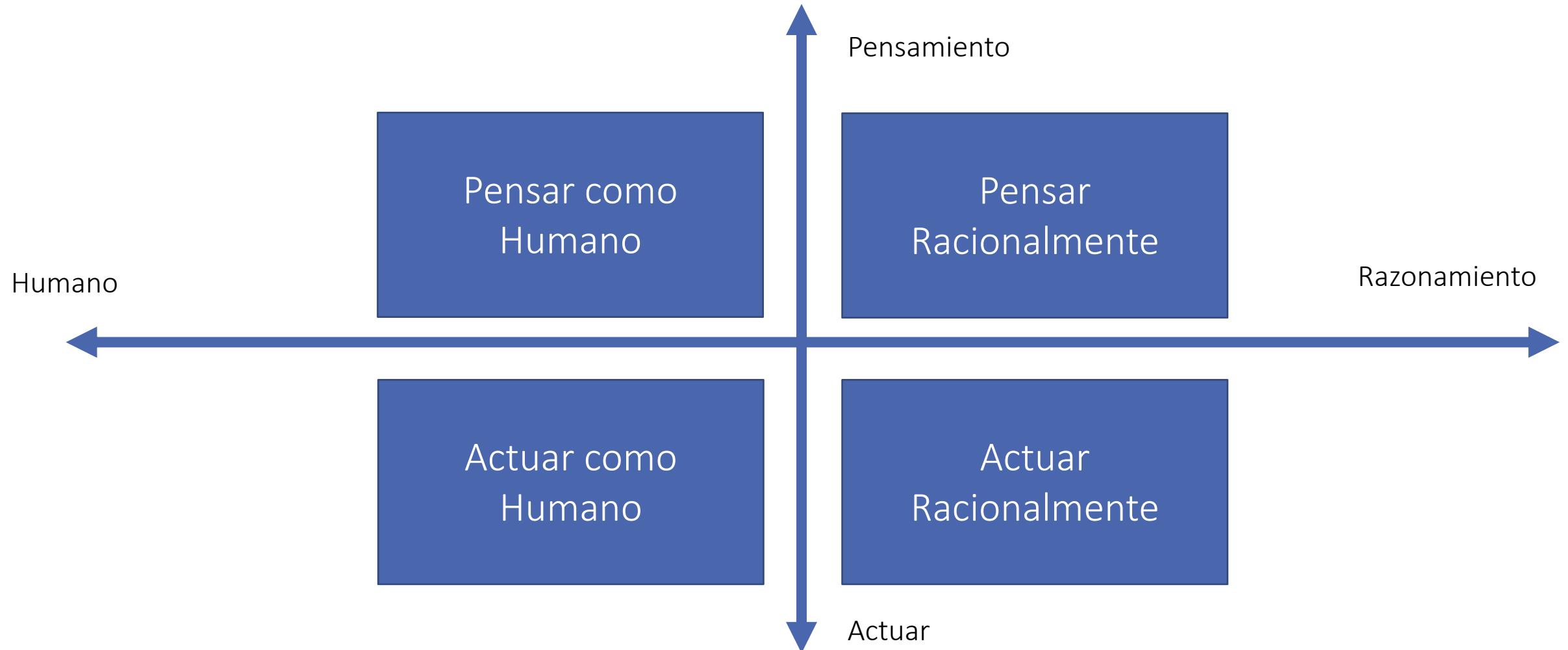
Inteligencia Artificial

¿Cómo describimos el comportamiento de algo inteligente?

¿Qué significa ser inteligente?

¿Cómo los humanos producen inteligencia?

Dimensiones de la Inteligencia Artificial



Pensar como Humano

Modelar la cognitiva humana:

En 1960 ocurre la revolución cognitiva, se basa en el estudio de la actividad cerebral de los humanos basado en lógica proposicional.

En 1957 Newell & Simon, conciben la idea de crear un sistema computacional capaz de resolver cualquier problema proporcionandole una descripción del problema, fue llamado General Problem Solver.

Pensar Racionalmente

Formalizar el proceso de construir inferencia:

Se desarrollaron varias notaciones para derivar reglas de pensamiento lógico, sirviendo como base para formar las bases matemáticas y lógicas de la inteligencia artificial moderna.

Promueve las bases sobre la lógica inferencial.

1) “*Todas las computadoras utilizan energía*”, 2) “*La energía produce calor*”, de este modo podemos concluir que “*todas las computadoras producen calor*”.

Actuar Racionalmente

Hacer lo correcto:

Se refiere a realizar lo que se espera como mejor solución en base a la información que se tiene disponible del problema.

Esto no necesariamente implica pensar. “reflejos humanos”

Desarrollar este enfoque en ambientes dinámicos puede ser muy complejo.

Actuar como Humano

Actuar como humano:

Crear sistemas que actúan como los humanos, para esto es necesario desarrollar varias capacidades:

Procesamiento de lenguaje natural,

Representar el conocimiento,

Desarrollar la capacidad de aprender,

Visión,

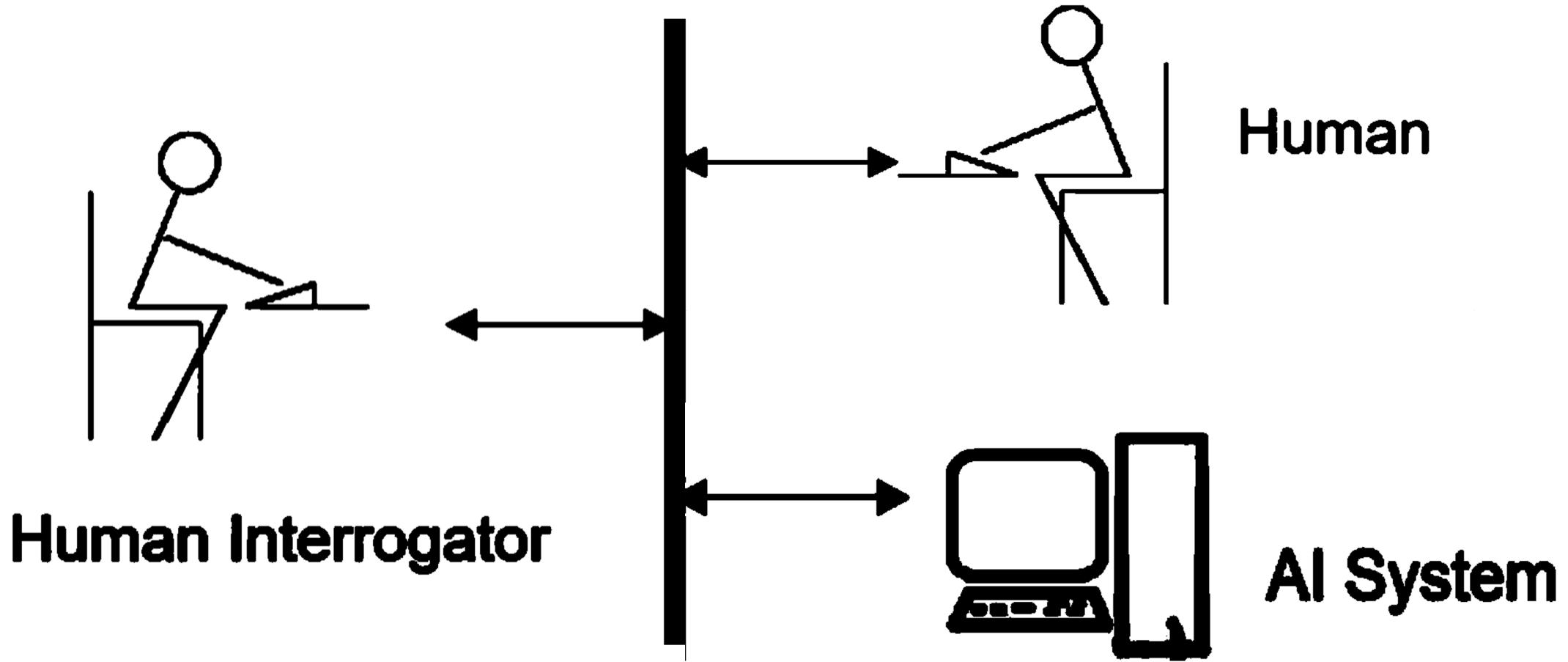
Mecanización (Robótica).

La prueba de Turing

En 1950, Alan Turing desarrolla la famosa prueba de Turing la cual establece una base formal para describir cuales son los componentes de una máquina verdaderamente inteligente.



La Prueba de Turing



La Prueba de Turing

Turing desarrollo mecanismos que permitían desifrar los códigos Nazis durante la segunda guerra mundial,

Estos dispositivos empleaban, Búsqueda Heurística, la cual formó parte de las ideas centrales sobre inteligencia artificial actual.

El Término “Inteligencia Artificial”

En 1956, 10 hombres del desarrollo del área de inteligencia artificialde aquellos días se reunieron durante 2 meses en Dartmouth College (Hanover, New Hampshire) a discutir cuales eran los aspectos más importantes sobre la inteligencia artificial, desde entonces se conoce el termino inteligencia artificial.

Esta reunión fue liderada por John McCarthy considerado uno de los padres de la inteligencia artificial.

Dicha reunión fue llamada “The Dartmouth Summer Research Project on Artificial Intelligence”

Inteligencia Debil vrs Inteligencia Fuerte

La inteligencia artificial debil: se centra en construir sistemas que actuan intelligentemente, tomando en cuenta que el sistema puede o no ser inteligente.

La inteligencia artificial fuerte: se centra en construir sistemas que son inteligentes (como las personas) esto implica no imitar la inteligencia. Es decir “sistemas con mentes”.

Inteligencia Artificial

“El objetivo de la inteligencia artificial es construir sistemas que resuelvan problemas los cuales requieren inteligencia humana”

Nilsson Nils (1971)

Esta definición no menciona explícitamente si se imita la forma en la que los humanos resuelven los problemas, únicamente habla sobre resolver los problemas que los humanos resuelven.

Sistemas Inteligentes: Implicaciones

Interacción con el mundo real: mecanización, reconocimiento del habla, procesamiento de imágenes,

Razonamiento y Planeación: representar el mundo real para tomar decisiones que resuelvan el problema lidiando con la incertidumbre de entorno.

Aprendizaje y Adaptación: Utilizar modelos de representación de aprendizaje para adaptarse a los cambios que puedan darse en el entorno del sistema.

¿Qué involucra?

Filosofía: Logica, métodos de razonamiento, Representación del conocimiento, fundamentos de aprendizaje.

Matemáticas: Representació formal y prueba de teorías, computación y algoritmos.

Estadística y Probabilidad: Modelado de la incertidumbre, aprendizaje sobre datos.

Economía: Funciones de utilidad y teoría de toma de decisiones.

¿Qué involucra?

Neurociencia: Modelado de elementos (neuronas) que permiten el procesamiento de información aglomerada (deep learning).

Psicología y Ciencia Cognitiva: Comprender cómo se comportan los humanos, percepción y procesamiento de sistemas cognitivos y representación del conocimiento.

Ingeniería de Computadoras: Construcción de computadoras con mejores prestaciones.

Teoría de Control: Diseñar sistemas que busquen maximizar una función objetivo conforme avance el tiempo.

Lenguística: Representación de la comunicación, gramática.

Revisión Histótica

1943: Inicios

McCulloch & Pitts: Constuyen un modelo lógico del cerebro humano, basado en el modelo neuronal.

1950's:

Turing publica su paper sobre la prueba de inteligencia,
Samuels Checkers program:

Primeros pasos de Machine Learning,

GPS:

Newell & Simons (Logicos teóricos).

Revisión Histórica

1943: Primeros Inicios

McCulloch & Pitts: Boolean Circuit Model Brain.

1950s:

Se crea la prueba de Turing.

Samuels: Crea el primer programa para jugar ajedrez.

Primeros Programas: Newell & Simons, Logic Theorist.

Revisión Histórica

1955 – 1965:

En la Dartmouth meeting se acuña el termino inteligencia artificial.

Nell and Simon: GPS, general problem solver,

McCarthy: Crea LISP.

1966 – 1973:

Se resuelven varios problemas de IA planteados previamente,

Se identifican algunas de las limitaciones que poseen los algoritmos de redes neurales.

Revisión Histórica

1969 – 1985:

Se crea la inteligencia artificial basada en conocimiento.

Los sistemas basados en reglas como Sistemas Expertos tiene mayor auge.

1986:

Crecimiento de Machine Learning,

Las Redes Neurales se vuelven populares nuevamente,

Se crean nuevos algoritmos de ML,

Se crean nuevos problemas de ML,

Revisión Histórica

Inicios de los 1990's

Se inicia a hablar sobre sistemas bajo incertidumbre.

Surge la idea de Redes Bayesianas.

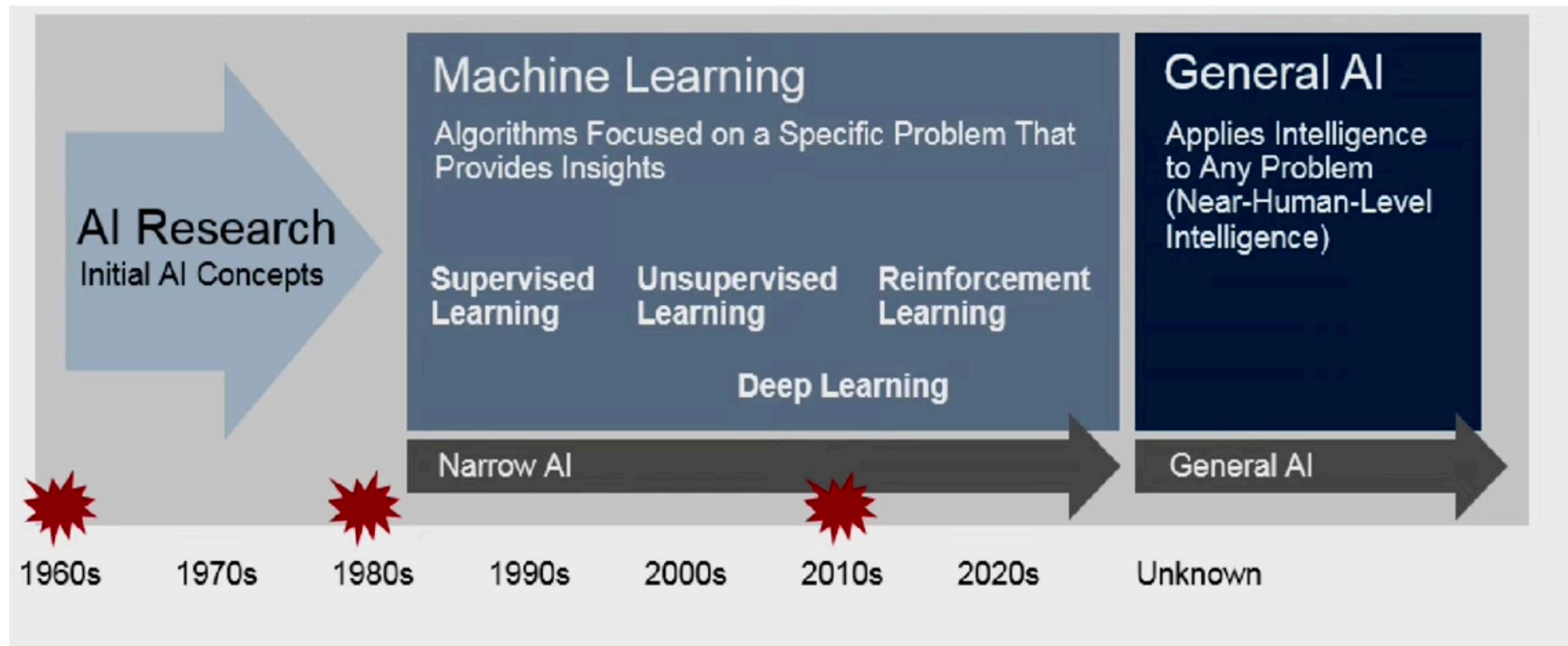
Inicios de 1995:

Se declara la Inteligencia Artificial como Ciencia,

Se integra el aprendizaje, resonamiento y representación del conocimiento como parte de un sistema inteligente.

Se inician a aplicar la IA en ramas como la visión, el lenguaje, minería de datos entre otras.

Revisión Histórica

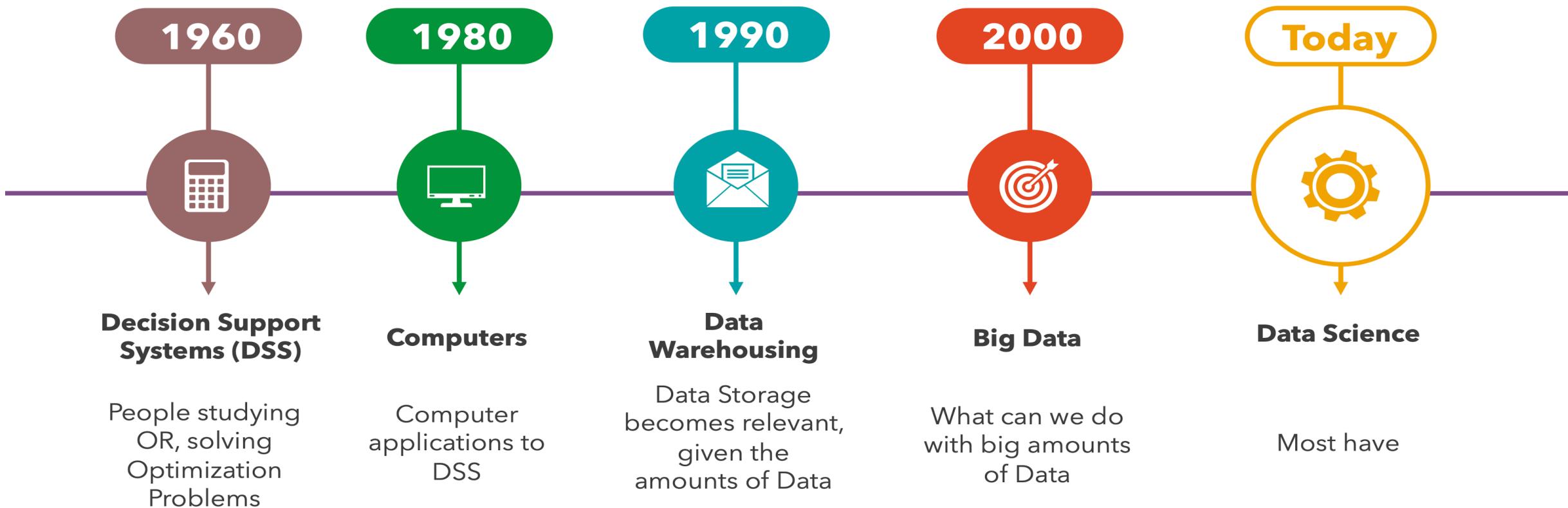


Lo de Hoy

- Modelos Convolucionales.
- Modelos secuenciales (Redes Neurales Recurrentes).
- Aprendizaje por Refuerzo.
- Aprendizaje Profundo por Refuerzo.

Evolución de Data Science

Evolución de Data Science





D. S.



Gran Capacidad de Procesamiento.



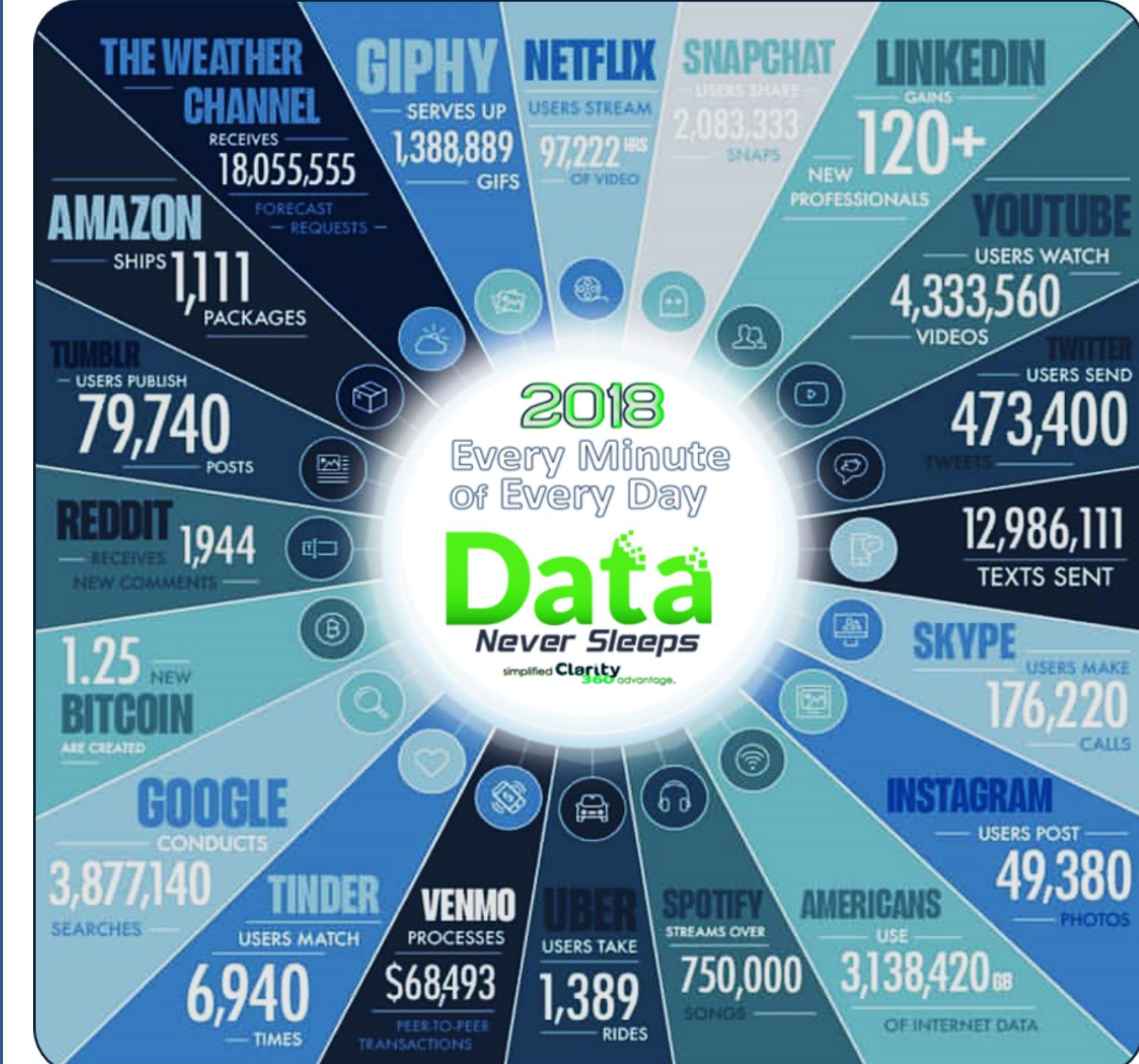
Grandes Cantidad de Información



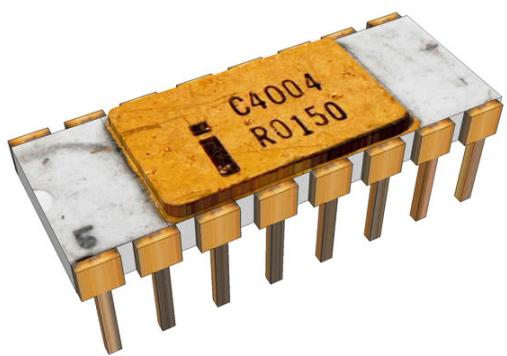
Algoritmos y Estructuras matemáticas

La Era de la Información

Almacenamiento



Procesamiento



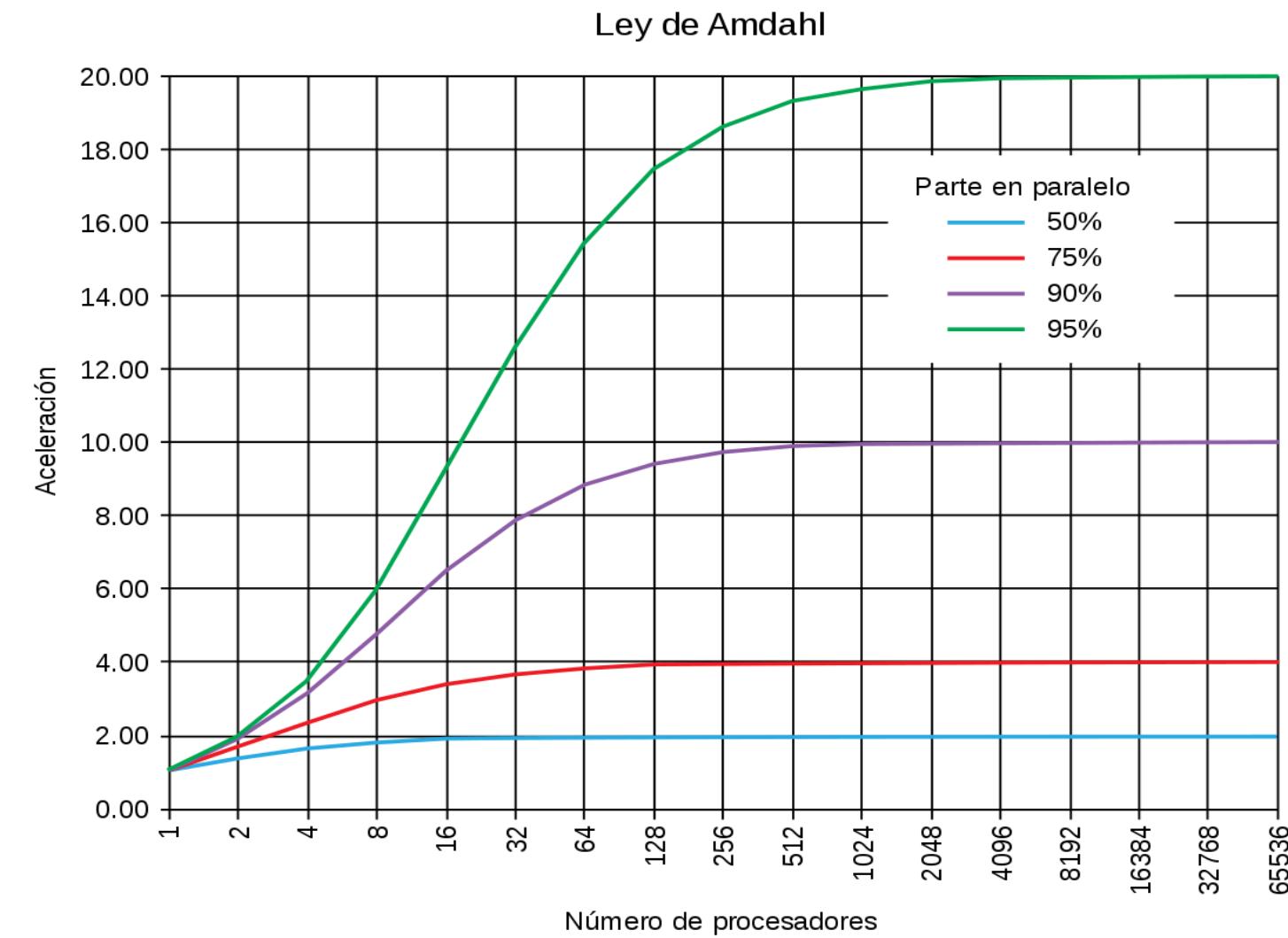
Ley de Amdahl: No todo es Color de Rosa

"A medida incrementamos la cantidad de procesadores en paralelo, la mejora en desempeño (speedup) decrece"

$$T(n) = T(1) \left((1 - p) + \frac{p}{n} \right)$$

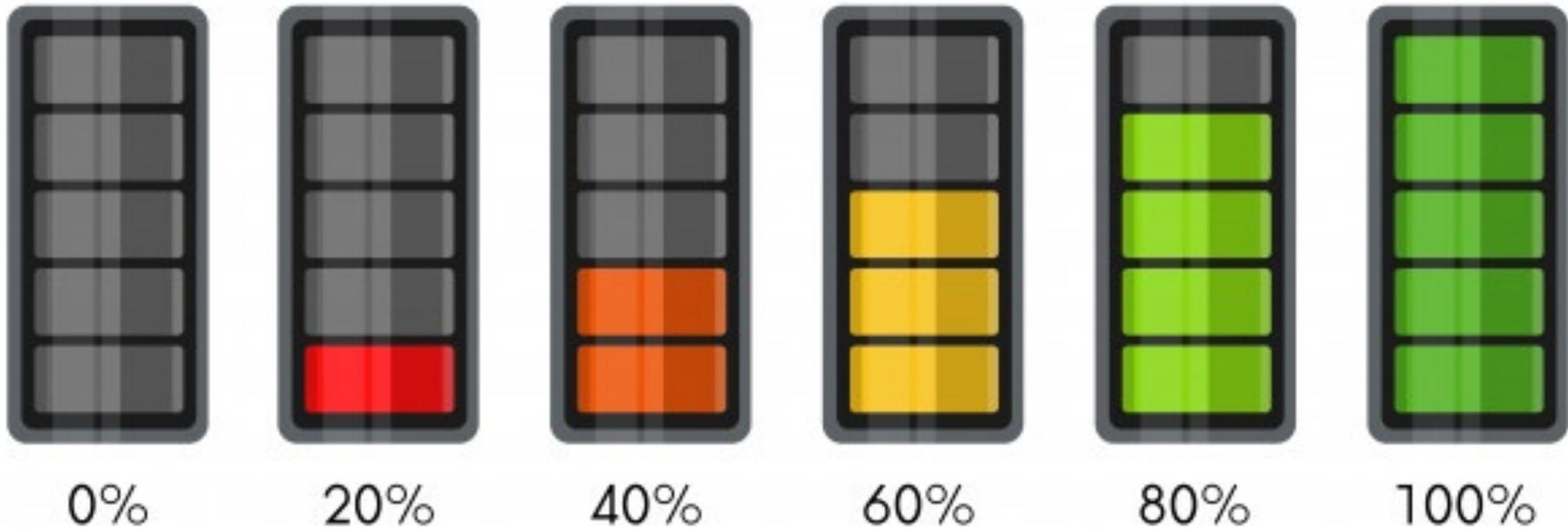
$$S(n) = \frac{T(1)}{T(n)} = \frac{1}{\left((1 - p) + \frac{p}{n} \right)}$$

Ley de Amdahl

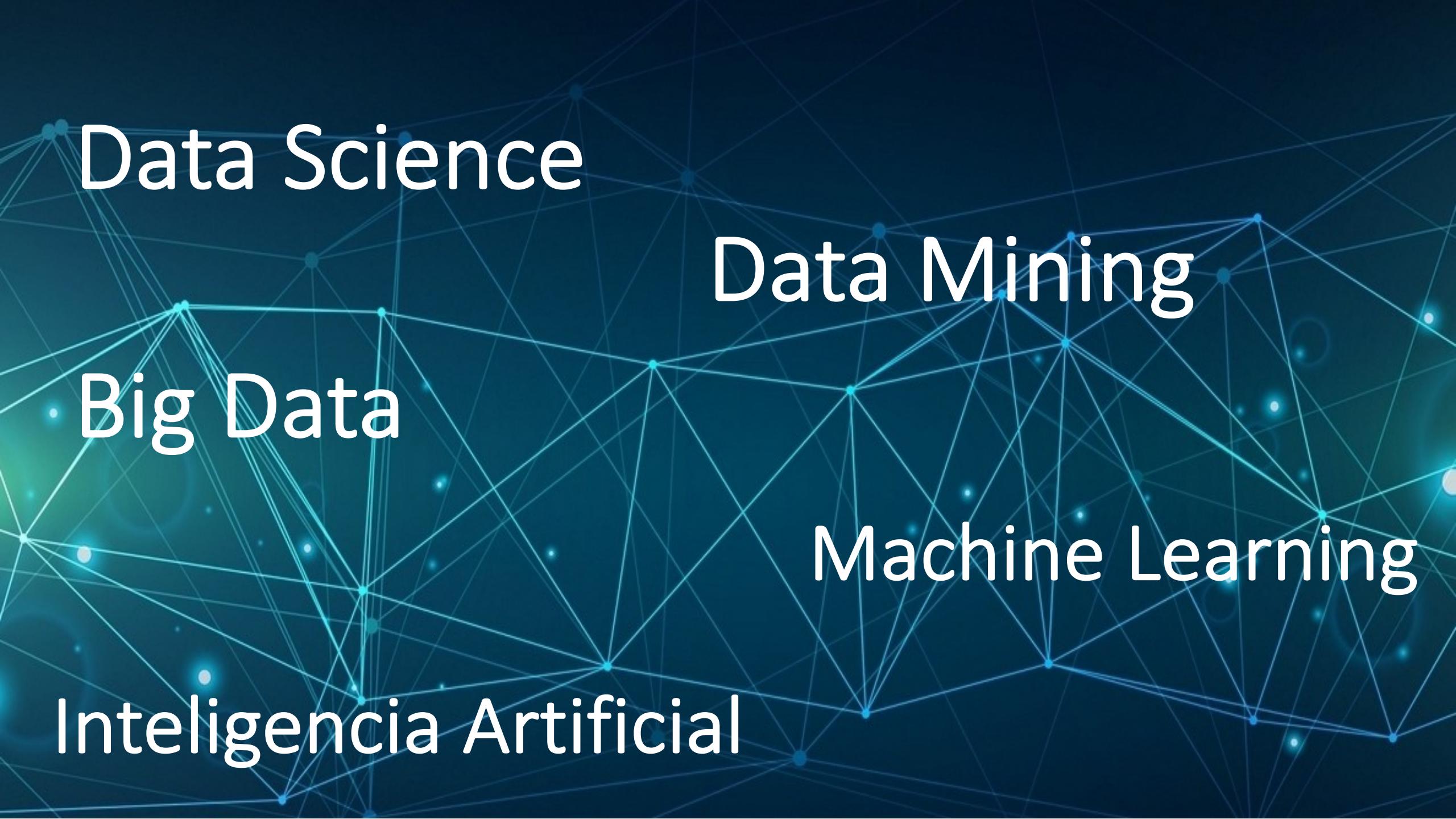


$$S(n) = \frac{1}{\left((1 - p) + \frac{p}{n} \right)}$$

¿El problema de hoy?



Algo de Terminología

The background of the image features a complex network graph. It consists of numerous small, semi-transparent blue dots of varying sizes scattered across the dark blue background. These dots are connected by a web of thin, light blue lines, creating a sense of a vast, interconnected system. Some clusters of dots are more densely packed than others, suggesting different levels of connectivity or data density.

Data Science

Data Mining

Big Data

Machine Learning

Inteligencia Artificial

● data science

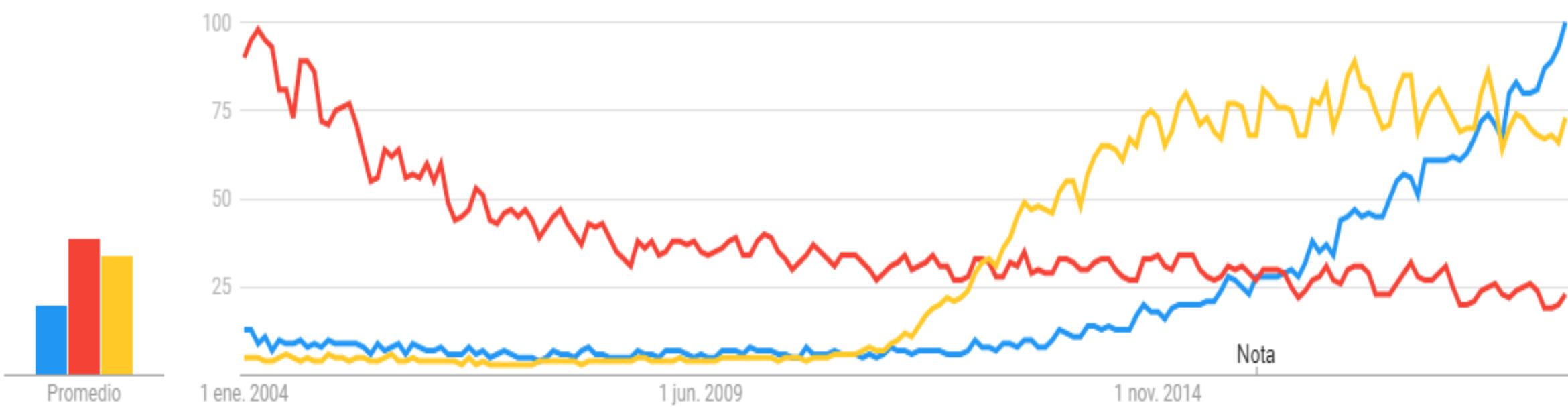
Término de búsqueda

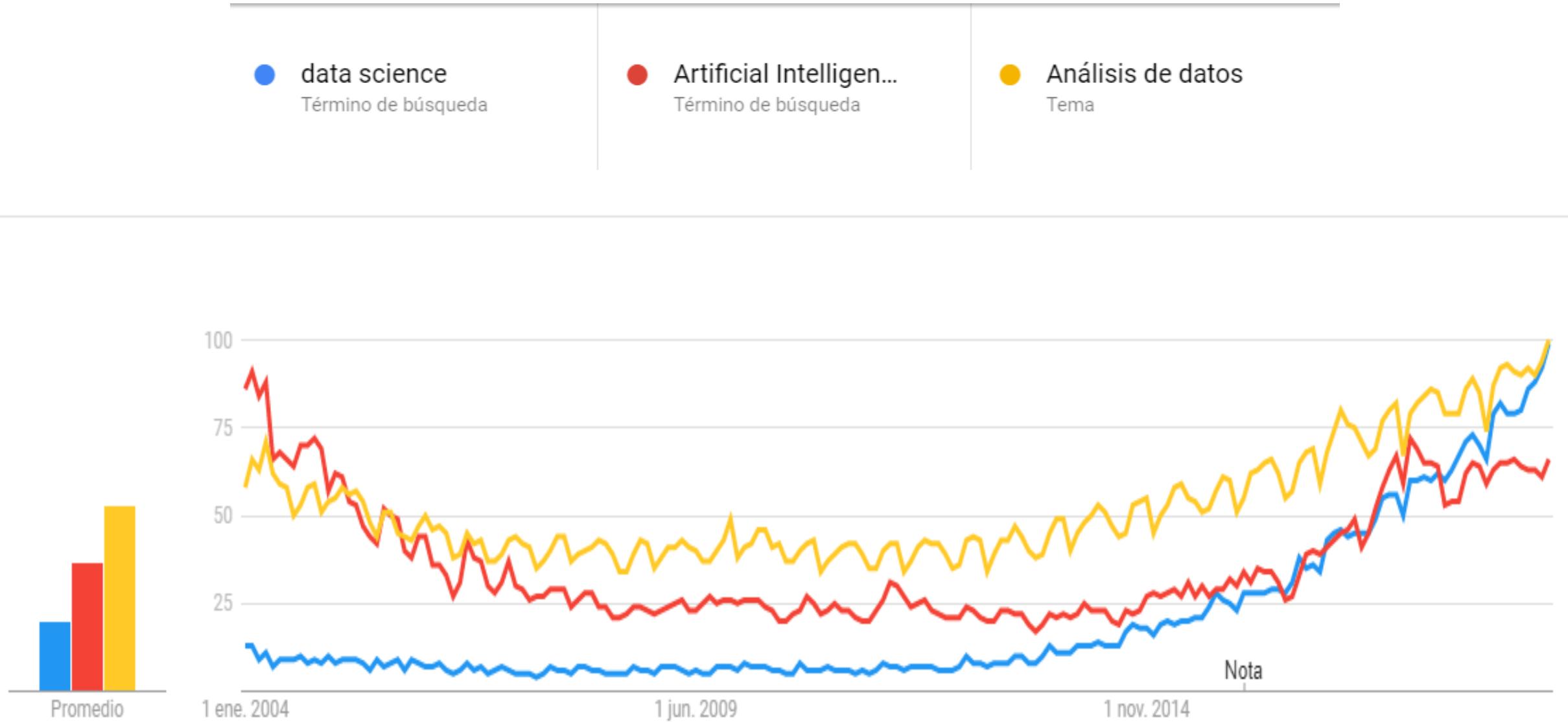
● Data Mining

Término de búsqueda

● Big Data

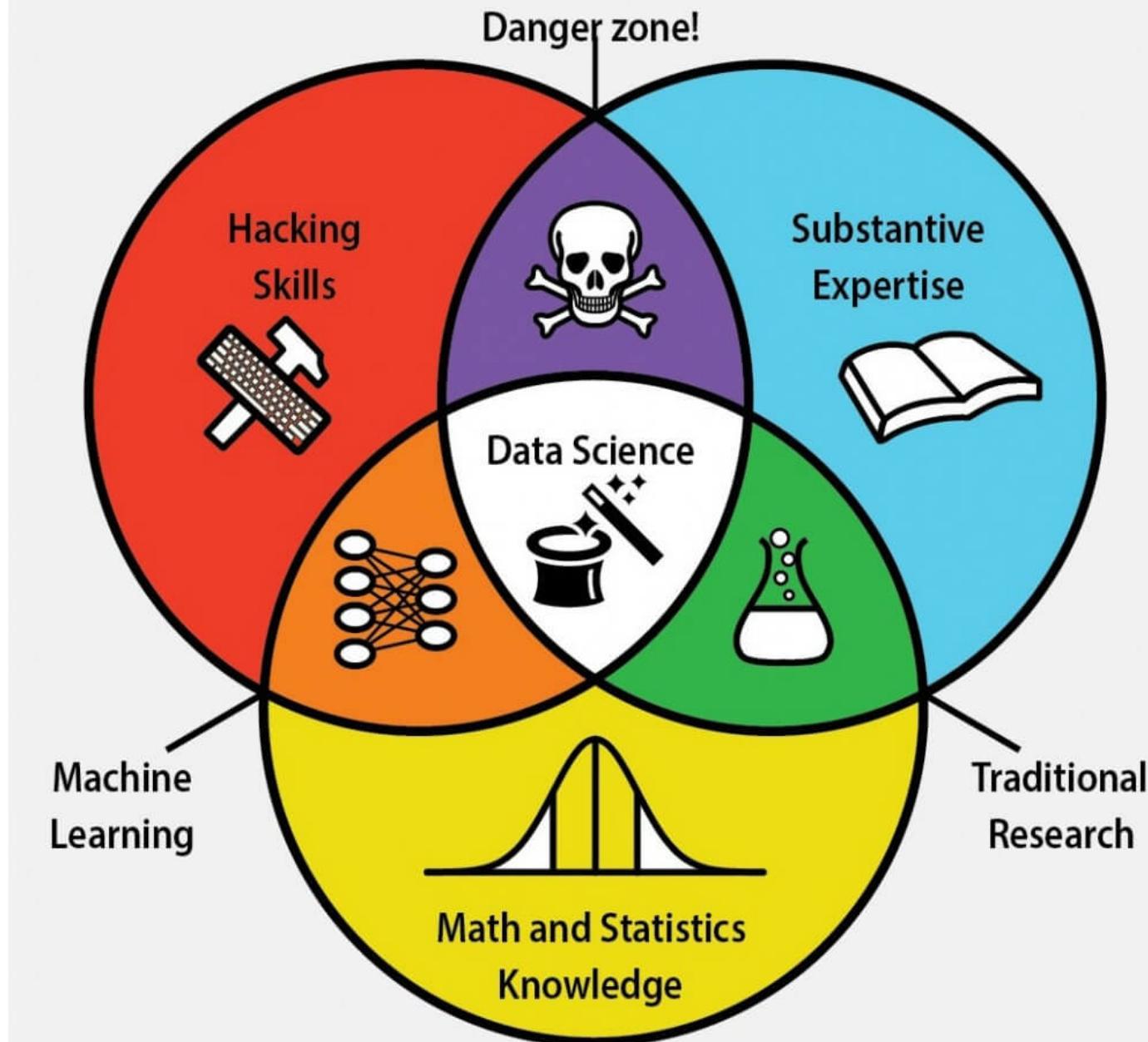
Término de búsqueda



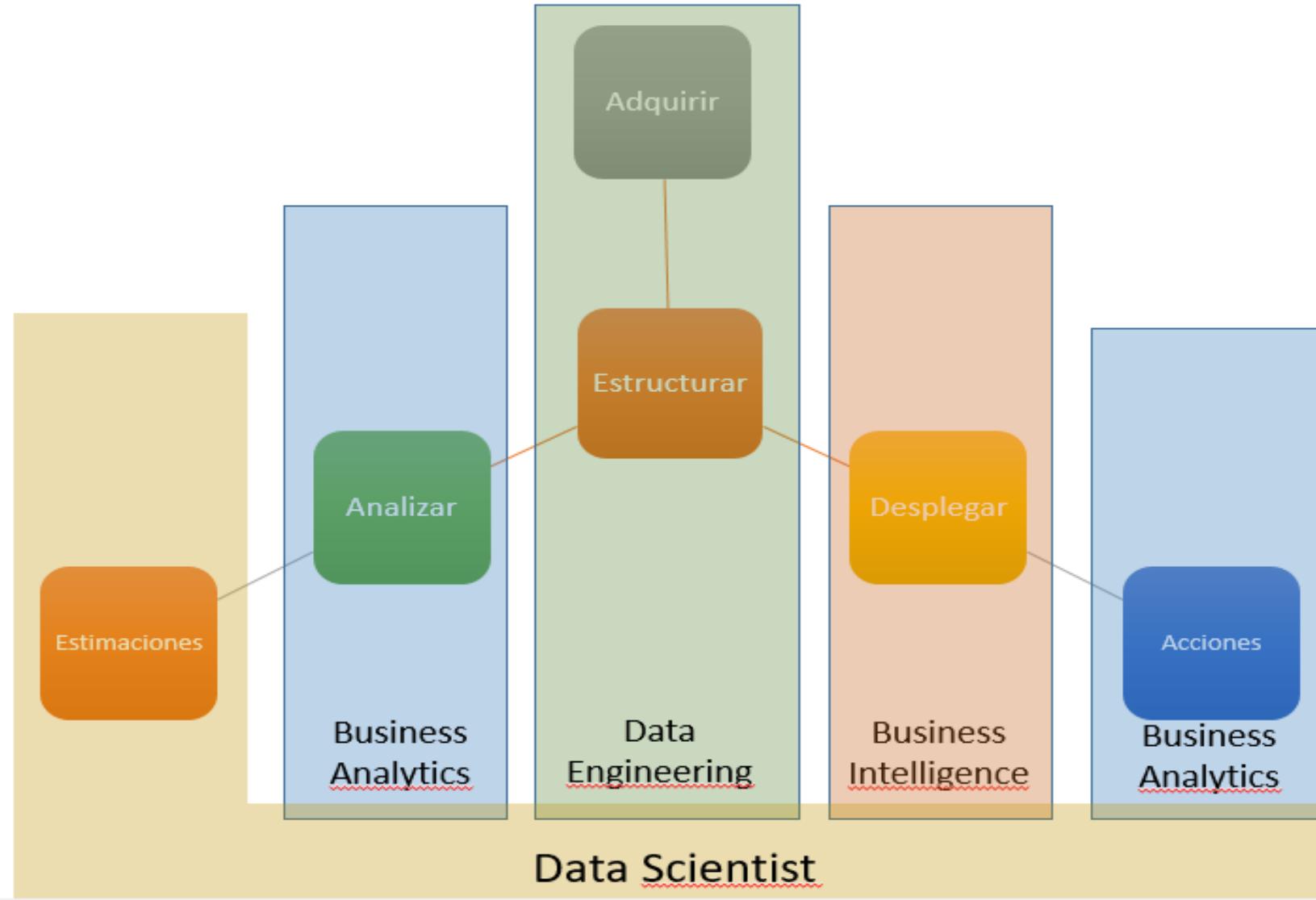


¿Qué necesita saber un
Científico de Datos?

DATA SCIENCE SKILLSET



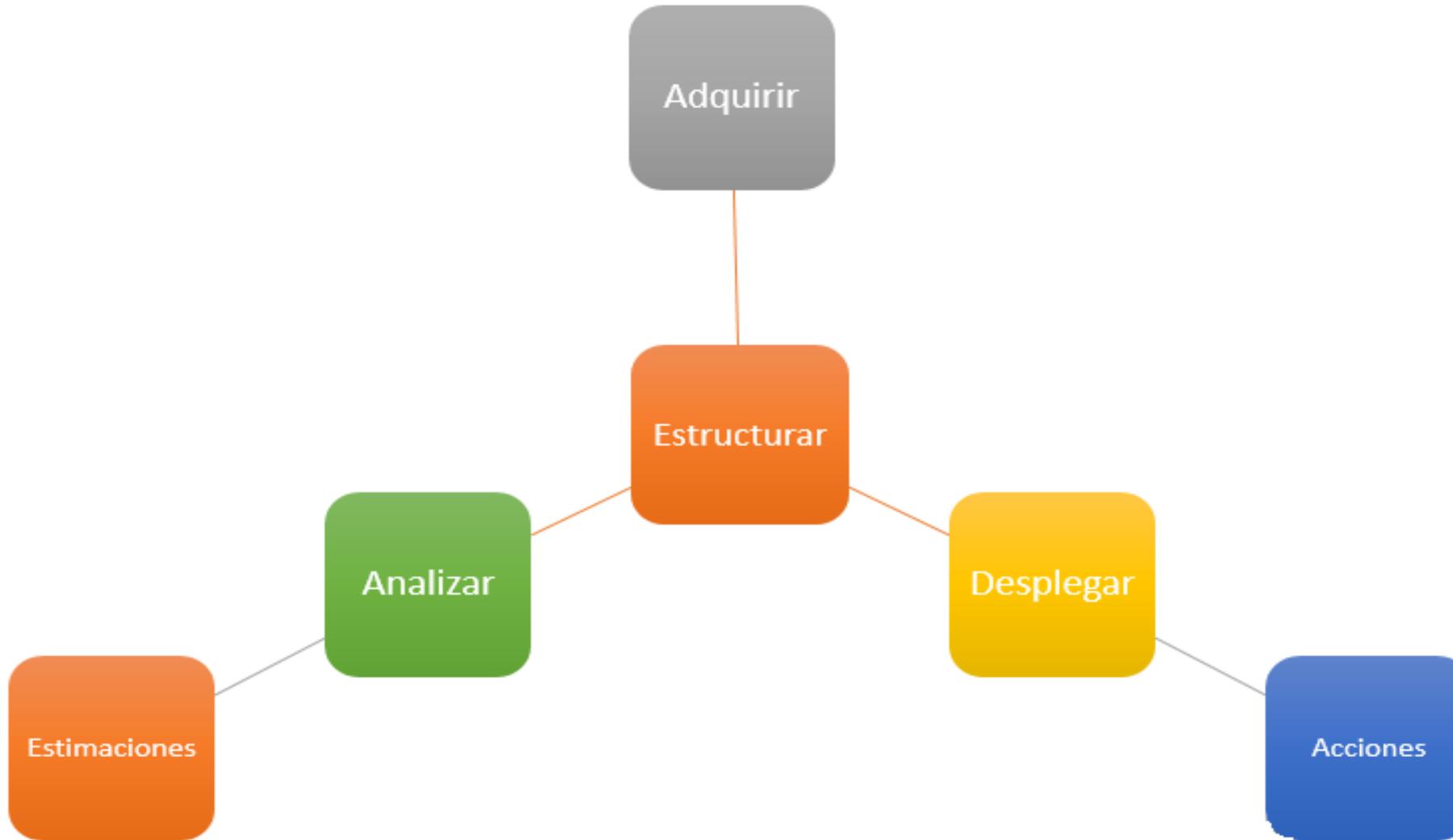
Roles del Científico de Datos



Algunos Conocimientos:

Data Engineering	Business Intelligence	Business Analytics	Data Scientist
<ul style="list-style-type: none">• Bases de Datos Avanzado• Programación Avanzada• Modelación Estadística y Matemática Medio• Múltiples Sistemas Operativos	<ul style="list-style-type: none">• Bases de Datos Medio• Programación Avanzada• Estadística y Matemática Medio• Dashboards	<ul style="list-style-type: none">• Administración• Economía y Finanzas• Modelación Estadística y Matemática Medio• Programación básica	<ul style="list-style-type: none">• Modelación matemática Avanzada• Estadística Avanzada• Programación avanzada• Bases de Datos Medio• Investigación

Tareas del Científico de Datos



Tareas del Científico de Datos

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative



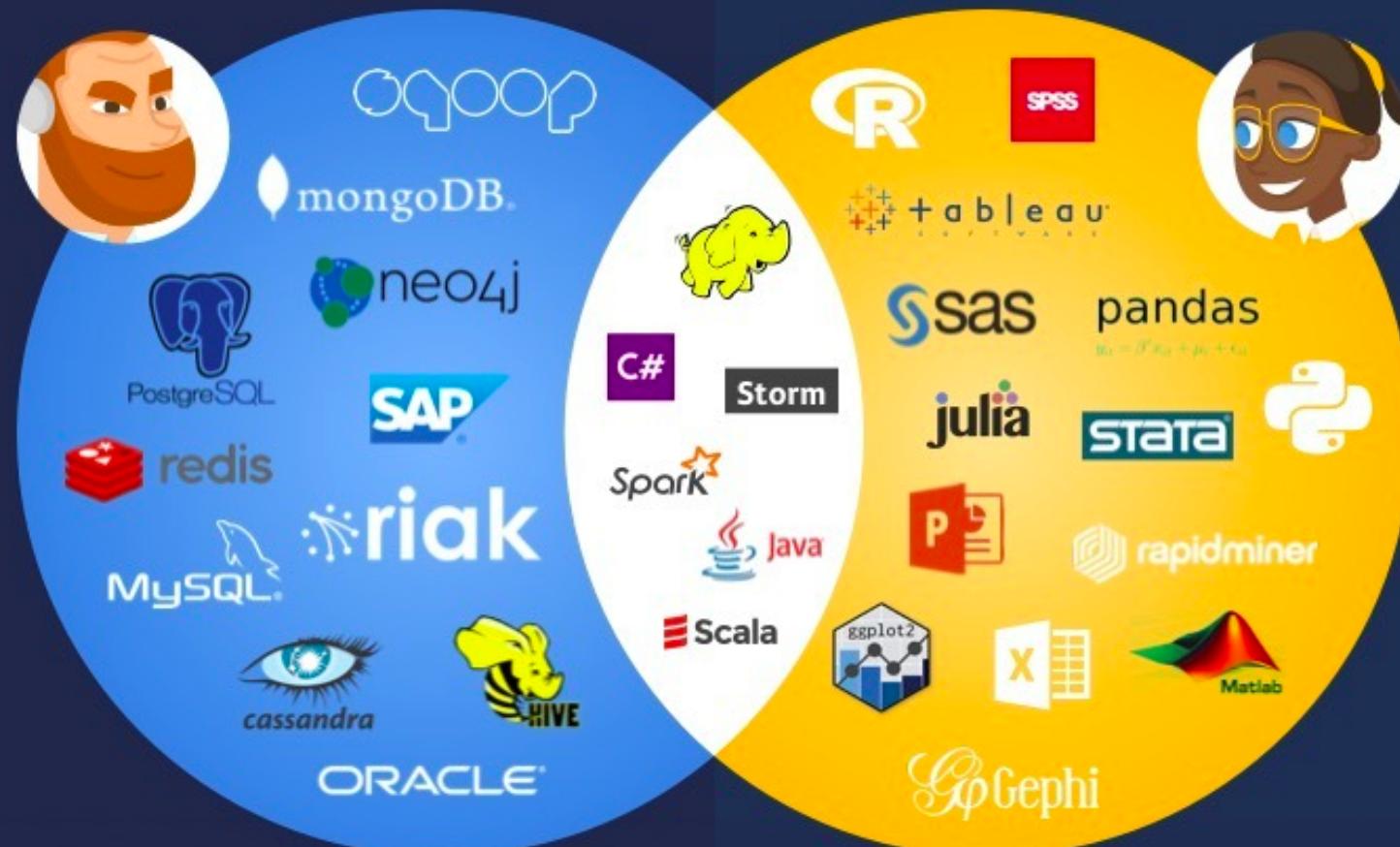
PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing package e.g. R
- ★ Databases SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data-driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

Languages, Tools & Software



Aplicaciones y Nuevas
Áreas.

SIGNET[®] BANK

5291 1512

6001 CP

VALID
THRU 02/93







TARGET

Sub áreas

- Minería de Texto.
- Procesamiento de Lenguaje Natural.
- Procedimiento de imágenes.
- Visión Artificial.
- Detección de Anomalías.

Nuevas Áreas

- Medicina Personalizada.
- Deep-Reinforcement learning.
- Medical Imaging.
- Healthcare.
- Knowledge Generation.