

Formación Profesional en CePETel 2022

Desde la Secretaría Técnica del Sindicato CePETel convocamos a participar del siguiente curso de formación profesional:

Big Data & Analytics – Parte 1

Clases: 8 de 3hs c/u de 18:00 a 21:00 hs.

Días que se cursa: los días miércoles 3, 10, 17, 24 y 31 de agosto; 7, 14, y 21 de septiembre.

Modalidad: a distancia (requiere conectarse a la plataforma Zoom en los días y horarios indicados precedentemente).

Docente: María Trinidad Aquino y Raúl Alejandro Grassi

La capacitación es:

- Sin cargo para afiliados y su grupo familiar directo.
- Sin cargo para encuadrados con convenio CePETel.
- Con cargo al universo no contemplado en los anteriores.

Informes: enviar correo a tecnico@cepotel.org.ar

Inscripción (hasta el 2 de agosto): ingresar al formulario (se recomienda realizar el registro por medio de una cuenta de correo personal y **no utilizar dispositivos de la empresa para acceder al link**).

<https://forms.gle/k64BVt8zouEP6E3u9>

Temario:

Módulo 1 Surgimiento de Big Data (Clases 1 y 2)

- Surgimiento de Big Data
- La revolución de los datos masivos, la datificación. Nuevo paradigma con cambio de escala.
- El Big Bang de Big Data. Razones del crecimiento explosivo de datos. Evolución de costo de almacenamiento.
- ¿Qué es Big Data? ¿Para qué sirve?
- La Paradoja de los Datos. TTTI (Total Time to Insight)
- Roles y Perfiles en Big Data

TP Clase 1: Ejercicio de búsqueda: casos posibles de uso de Big Data

TP Clase 2: Detallar 3 características diferenciales de las bases de datos relacionales y no relacionales.

Ing. Daniel Herrero – Secretario Técnico – CDC

Módulo 2 Los datos nos rodean (Clases 3 y 4)

- Implementación de Big Data
 - ¿Qué es el procesamiento distribuido?
 - Diferencias entre las bases de datos existentes: SQL/NoSQL
 - Aproximación al Ecosistema Hadoop.
 - La relación entre BigData y la nube: Cloud vs On Premise
 - Ejemplos de aplicaciones de Big Data en distintos ámbitos – casos reales de uso
- TP Clase 3: Identificar y describir al menos 3 diferencias entre los servicios de cloud vs on premise.
TP Clase 4: Diferenciación entre Inteligencia Artificial, Machine Learning y Deep Learning.
Ejemplos

Módulo 3 Obtención de los Datos y Proceso de Toma de Decisión (Clases 5 y 6)

- Inteligencia Artificial • Machine Learning y sus algoritmos
 - Deep Learning • Definición y Tipos de Data Analytics
 - Business Intelligence: Datos vs Información vs Conocimiento
 - BI: tipos de sistemas. OLTP vs OLAP
 - Desafíos de BI. La pirámide de la información
- TP Clase 5: Marcar la diferencia entre los sistemas OLTP y OLAP. Traer un ejemplo de al menos 1 de ellos para compartir en la próxima clase.
TP Clase 6: Listar al menos 2 herramientas para la visualización de datos. Comparación con ventajas y desventajas de cada una.

Módulo 4 Visualizar con datos (Clases 7 y 8)

- Los datos empiezan a hablar
 - Niveles de realización de BI: Alertas, Dashboards, Monitoreos
 - Análisis OLAP/Multidimensional
 - Herramientas de Explotación de BI
 - Gráfico cuadrangular de Gartner: clasificación
 - Herramientas para la visualización con datos masivos
 - KPI. Definición, Ejemplos. Diferencias entre KPI's y Métricas
- TP Clase 7: Buscar ejemplos de visualizaciones con las que trabajen o hayan trabajado. Armar una presentación de no más de 3 minutos por alumno o por grupo (a definir según los participantes). Ejemplos de KPI más comunes según la industria.

Acerca de los docentes

María Trinidad Aquino: 2007-12-03 al momento Analista Senior Marketing – Región Patagonia Movistar, Neuquén.

* Proyecto Canal Presencial (Agentes y CEC) responsable de la construcción y disposición de los datos (KPI's) para ver su evolución con aporte analítico, diagnóstico y sugerencia de planes de acción para su mejora.

Ing. Daniel Herrero – Secretario Técnico – CDC

*Proyecto Educador Digital País, referente de la región Patagonia con seguimiento de los KPI's y evolución.

* Nuevos proyectos del área con análisis de datos, participación en el diseño, elaboración y difusión de lo implementado.

Formación académica:

- 2020-09 – 2021-10 (finalizada) Big Data, Data Engineer (Diplomatura) ITBA, CABA
- 1999-03 - 2003-12 (finalizada) Ciencias Sociales, Licenciada en Relaciones Públicas Universidad Nacional de Lomas de Zamora, Lomas de Zamora

Raúl Alejandro Grassi: desde 1995 hasta la fecha TELEFONICA DE ARGENTINA S.A.
Puesto:Analista Senior - Sector Big Data Comercial.

Responsabilidades: Definición de inversiones anuales en capital (CapEx) en base a análisis de proyección comercial. Gestión de proyectos y seguimiento de inversiones. Diseño e implementación de modelos de aseguramiento de satisfacción de clientes. Planeamiento y ejecución de tableros de control y análisis del negocio basado en datawarehousing (heavy user) en los últimos 10 años, programando en SQL y modelado de datos. Análisis y Evaluación de acciones que impacten en cumplimiento de objetivos del Negocio B2C. Analista Senior BI, desarrollo en herramientas de explotación de BI (Microstrategy; Tibco Spotfire, Power BI, Tableau, etc.) y ecosistema Hadoop (Spark, Hive, SQL, procesos de ingestas ETL, ELT, etc.).

Desempeño durante 4 años en el sector Data Driven Comercial, promoviendo la cultura Data Driven y desarrollando tableros de control predictivos y prescriptivos con herramientas de explotación basadas en modelos relacionales/dimensionales.

Experiencia al menos 7 años como líder de proyectos, Manejo de Metodologías Agiles en

posiciones como Stakeholder, Scrum Master y PO.

Formación académica:

- 2020-2021 Licenciatura en Big Data – especialista en Data Engineer - ITBA (Instituto Tecnológico de Buenos Aires)
- 1999 Posgrado en Gestión Gerencial Avanzada (Management Executive Program) Universidad Argentina de la Empresa (UADE)
- 1986-1992 Ingeniero Electrónico Universidad de Buenos Aires

Ing. Daniel Herrero – Secretario Técnico – CDC

<http://www.cepetel.org.ar> ✉ tecnico@cepetel.org.ar 📍 Rocamora 4029 (CABA) ☎ (+54 11)35323201

BIG DATA & ANALYTICS I



TEMARIO

- Módulo 1: Surgimiento de Big Data
- Módulo 2: Los Datos Nos Rodean
- Módulo 3: Obtención de los Datos y Proceso de Toma de Decisión
- Módulo 4: Visualizar con Datos

Disertantes: Lic. Maria Trinidad Aquino – Ing. Raúl Alejandro Grassi

CePETel

Sindicato de los Profesionales
de las Telecomunicaciones

SECRETARÍA TÉCNICA



Instituto Profesional de
Estudios e Investigación

BIG DATA & ANALYTICS I

Módulo 1: Surgimiento de Big Data

SURGIMIENTO DE BIG DATA

1991

Surgimiento de Internet.

1996

El precio de [almacenamiento digital](#) cae. Es más rentable que el papel.

1997

[Google](#) lanza su motor de búsqueda.

1999

Primeras ideas de [Big Data](#)

El primer uso del término Big Data [en un trabajo académico](#): “Visually Exploring Gigabyte Datasets in Realtime (ACM)”.

El primer uso del concepto “[Internet of Things](#)” (Internet de las Cosas) en una presentación de negocios de Kevin Ashton para Procter and Gamble.

2001

Las tres “V” definida por Doug Laney para [Big Data](#): “Volume, Velocity, Variety”

2004

Se funda [Facebook](#) en Cambridge, Massachusetts.

2005

La [Web 2.0](#) aumenta el volumen de datos.

2006

Se funda [Twitter](#) en San Francisco, California.

2007

Emerge el uso actual del término “[Big Data](#)”

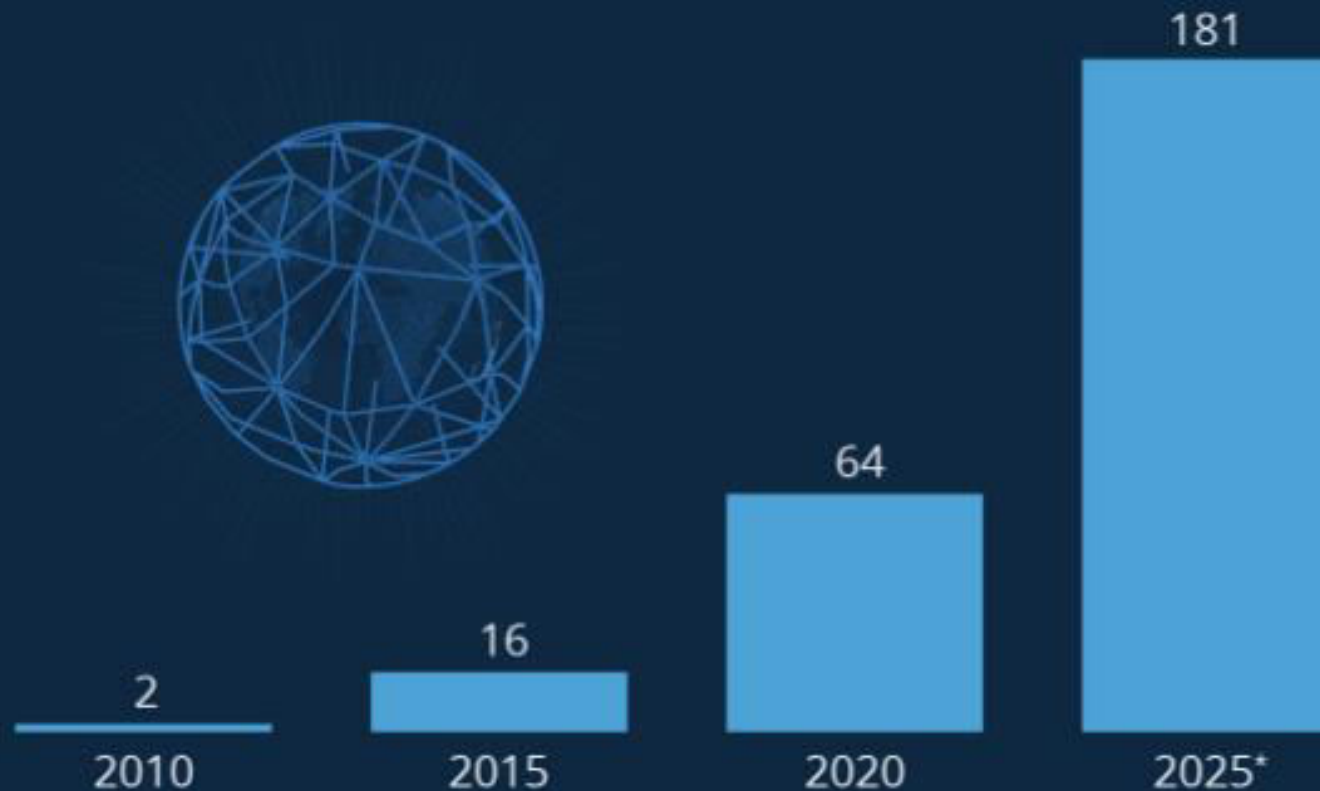
Steve Jobs, Apple, lanza el [Iphone](#).

[Video Bienvenida Cronología de Big Data](#)

- 2009 Una compañía promedio estadounidense con más de 1.000 empleados, almacena más de 200 terabytes de datos de acuerdo al informe sobre Big Data: “The Next Frontier for Innovation, Competition and Productivity” realizado por McKinsey Global Institute.
- 2010 Eric Schmidt, presidente ejecutivo de Google, dice en una conferencia que la cantidad de datos que ahora se están creando cada dos días, es mayor que la creada desde el comienzo de la civilización humana hasta el año 2003.
- 2014 El uso de Internet móvil supera a las computadoras de escritorio por primera vez. El 88% de los ejecutivos que respondieron a una encuesta internacional realizada por GE dice que el análisis de grandes volúmenes de datos es una prioridad.

El Big Bang del Big Data

Volumen estimado de datos digitales creados o replicados en todo el mundo, en zetabytes



Un zetabyte equivale a mil millones de gigabytes.

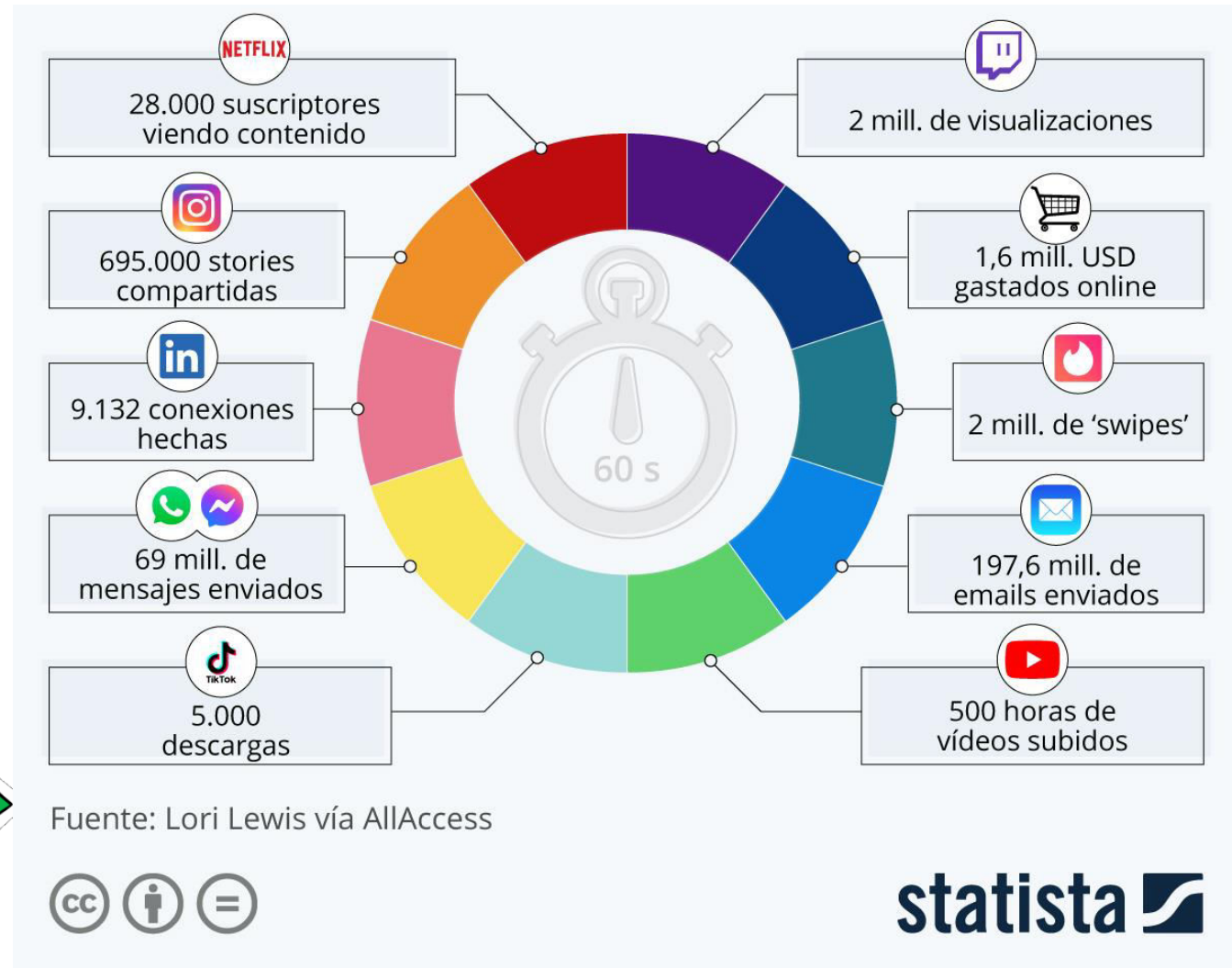
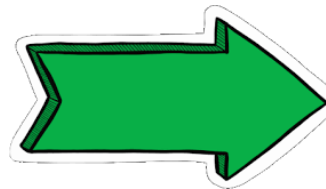
* Previsión de marzo de 2021.

Fuentes: IDC, Seagate, Statista

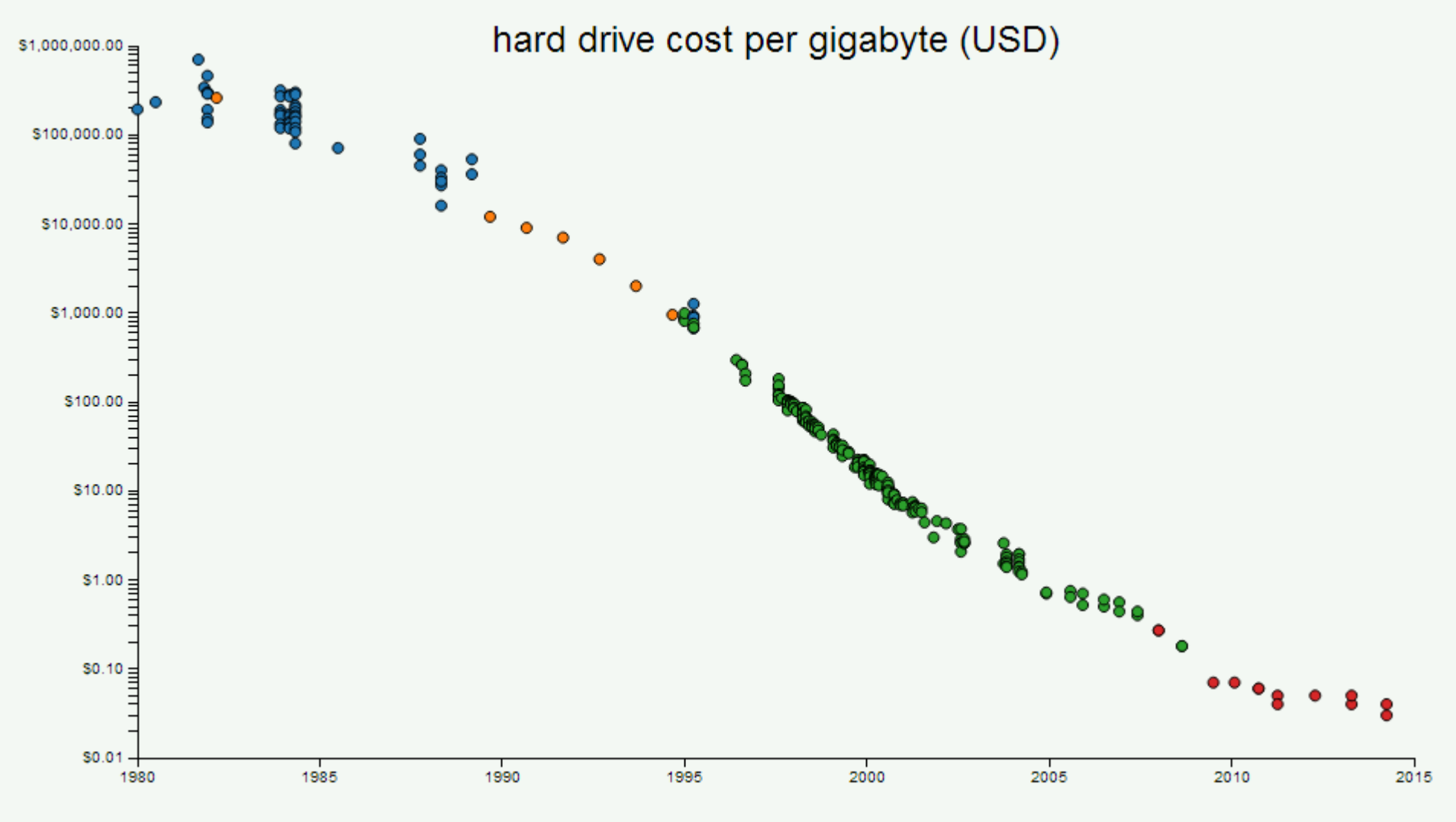
Los datos nunca duermen!

El mundo pasa una media de 6 horas y 54 minutos al día en Internet, pero, en sólo un minuto, ¿cuánta actividad digital se genera? 60 segundos pasan rápidamente, pero en este lapso de tiempo suceden un sinfín de cosas en el mundo online.

Esto sucede en internet en 1 minuto

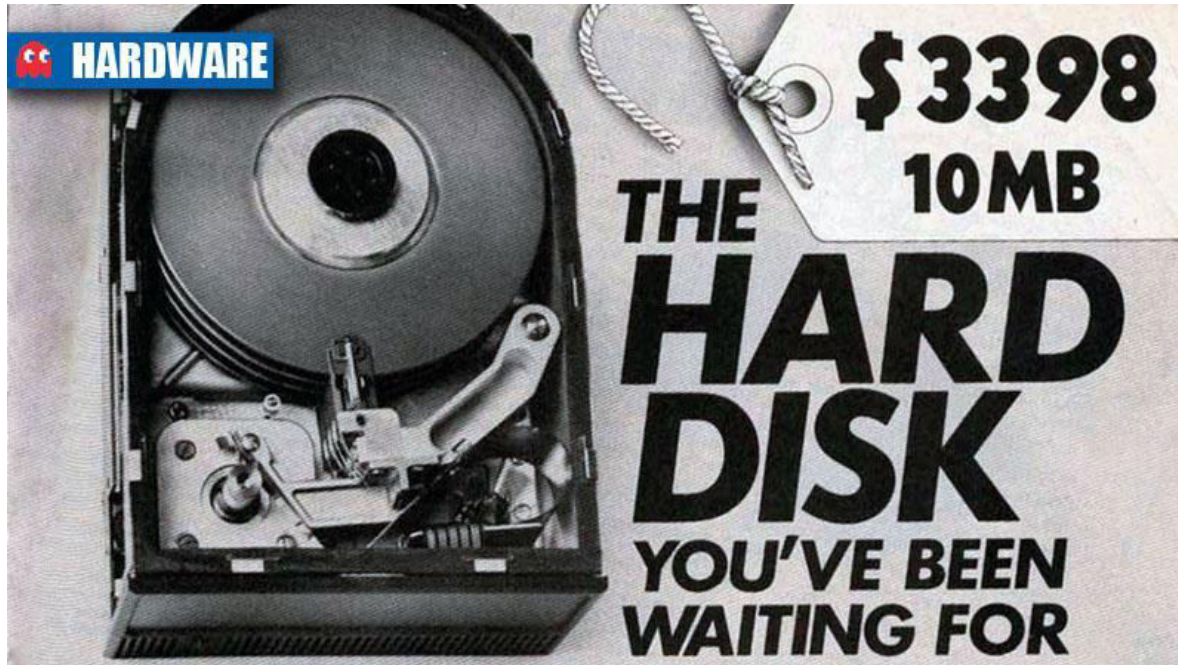


CRECIMIENTO EXPLOSIVO DE DATOS VS. COSTO DE ALMACENAMIENTO



CRECIMIENTO EXPLOSIVO DE DATOS VS. COSTO DE ALMACENAMIENTO

Año 1989



24 de Agosto de 1997		
Marca	Capacidad	Precio
Western Digital	2.1 gigabytes	u\$279.99
Western Digital	3.1 gigabytes	u\$329.99
Maxtor	3.5 gigabytes	u\$359.99
Maxtor	4.3 gigabytes	u\$439.99
Western Digital	5.1 gigabytes	u\$459.99

Alojamiento gratuito

¿Cuánto puedes almacenar gratis?

	OneDrive de Microsoft	Google Drive	iCloud Drive	Amazon Cloud Drive	Dropbox
A de junio 2015	15GB	15GB	5GB	5GB	2GB

Service:	Price per 1TB (yearly 1TB plan or closest)
iDrive	\$1.33 per month
Sync.com	\$4 per month
pCloud	\$4.17 per month
Icedrive	\$4.17 per month
Google Drive	\$4.17 per month
MEGA	\$4.75* per month
Dropbox	\$4.99 per month
OneDrive	\$5.83 per month
Koofr	\$11.80* per month

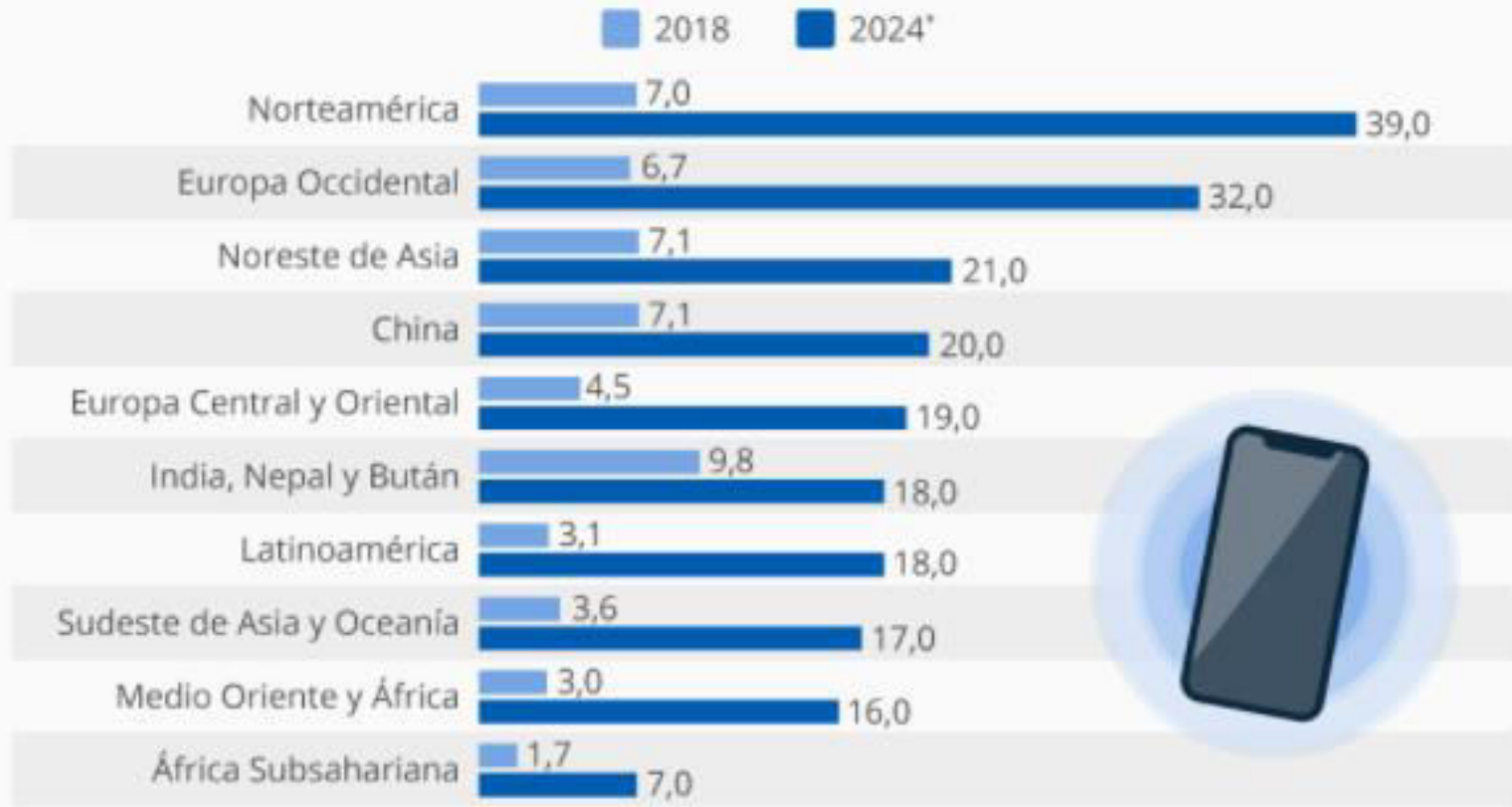
CRECIMIENTO EXPLOSIVO DE DATOS VS. COSTO DE ALMACENAMIENTO



CRECIMIENTO EXPLOSIVO DE DATOS VS. COSTO DE ALMACENAMIENTO

El uso de datos móviles no para de crecer

Volumen de tráfico de datos por smartphone (en gigabytes/mes)



Big Data

Una Primer Definición

“*Volumen masivo de datos*, tanto *estructurados como no-estructurados*, los cuales son *demasiado grandes y difíciles de procesar* con las bases de datos y el software *tradicionales*.”

(ONU, 2012)



Big Data no es sólo un repositorio de datos:

además de almacenar los datos, se aplican procesos para tratarlos y analizarlos.

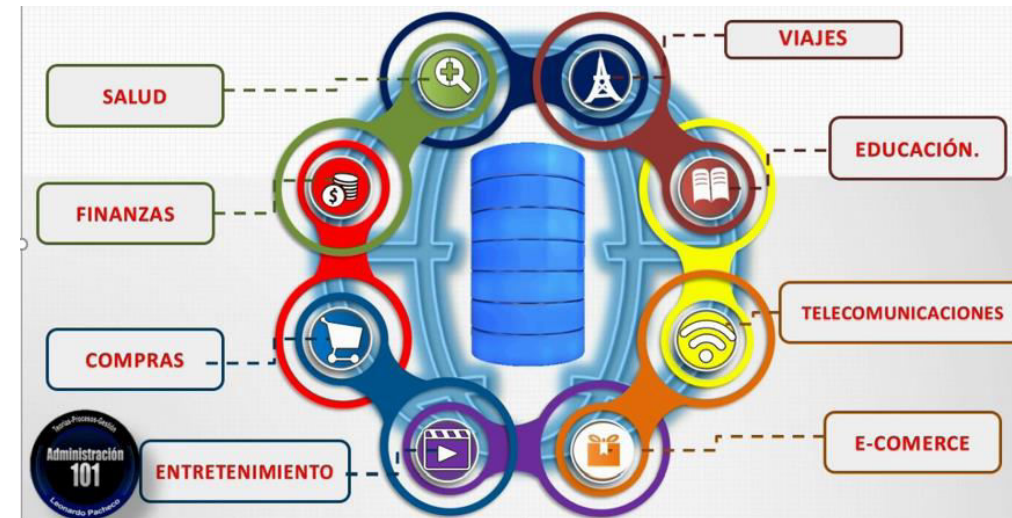
Big Data no es únicamente una herramienta de analítica: es capaz de correlacionar datos procedentes de distintas fuentes y en diferentes formatos.

Big Data es más que una plataforma de gestión de datos: es capaz de trabajar con datos en tiempo real.

Big Data

Para que sirve?

- **Predecir el consumo de energía**, muy útil para los servicios públicos
- **Anticipar la deserción de clientes**, por ejemplo, en compañías de telecomunicaciones
- **Analizar el sentimiento** hacia una marca en redes sociales, muy útil en marketing
- **Mejorar el rendimiento** de los sistemas de atención al cliente
- **Crear sistemas de recomendación automatizados**, con muchas aplicaciones en el sector retail
- **Detectar fraudes**, una aplicación que puede ser muy útil para aseguradoras y bancos, y también en recursos humanos



Big Data

nos ayuda a:

- Identificar personas, con la utilidad que supone para **cuestiones de seguridad y defensa**
- **Conocer mejor los gustos y necesidades de los clientes**, para ofrecerles productos y servicios personalizados
- **Facilitar el día a día de las personas con dispositivos inteligentes**: neveras que hacen el pedido directamente al supermercado, coches conectados que informan de que hay retenciones en una vía o pulseras para diabéticos que miden el nivel de glucosa a través de la piel.



ENTONCES
QUE ES EL
BIG DATA??



Big Data

Es el sector de IT que hace referencia a la gestión y análisis de **grandes conjuntos de datos** que por la **velocidad** a la que se generan, la capacidad para tratarlos y los **múltiples formatos y fuentes**, es necesario procesarlos con mecanismos distintos a los tradicionales.

CADA DÍA CREAMOS 2,5
QUINTILLONES DE BYTES DE
DATOS. (2,5 Exabytes)

EL 90% DE LOS DATOS DEL
MUNDO DE HOY SE
GENERARON EN LOS
ÚLTIMOS 2 AÑOS

[Video: Big Data en 3 minutos!!!](#)



De donde vienen los datos?



- Los **datos pueden llegar** desde las redes sociales, a través del correo electrónico, de las apps... **son datos que genera el usuario por sí mismo.**
- Además, por ejemplo, con las **transacciones del banco**, de un *e-commerce*, o de las tarjetas de crédito, estos datos **también quedan registrados.**
- Las **propias máquinas** también **generan datos por sí mismas**, y codifican información a través de **sensores.**
- También se van **archivando** los datos relacionados con **seguridad, defensa e inteligencia.** Un claro ejemplo de ello son los que generan los **lectores biométricos.**

Gestión de datos



Almacenarlos

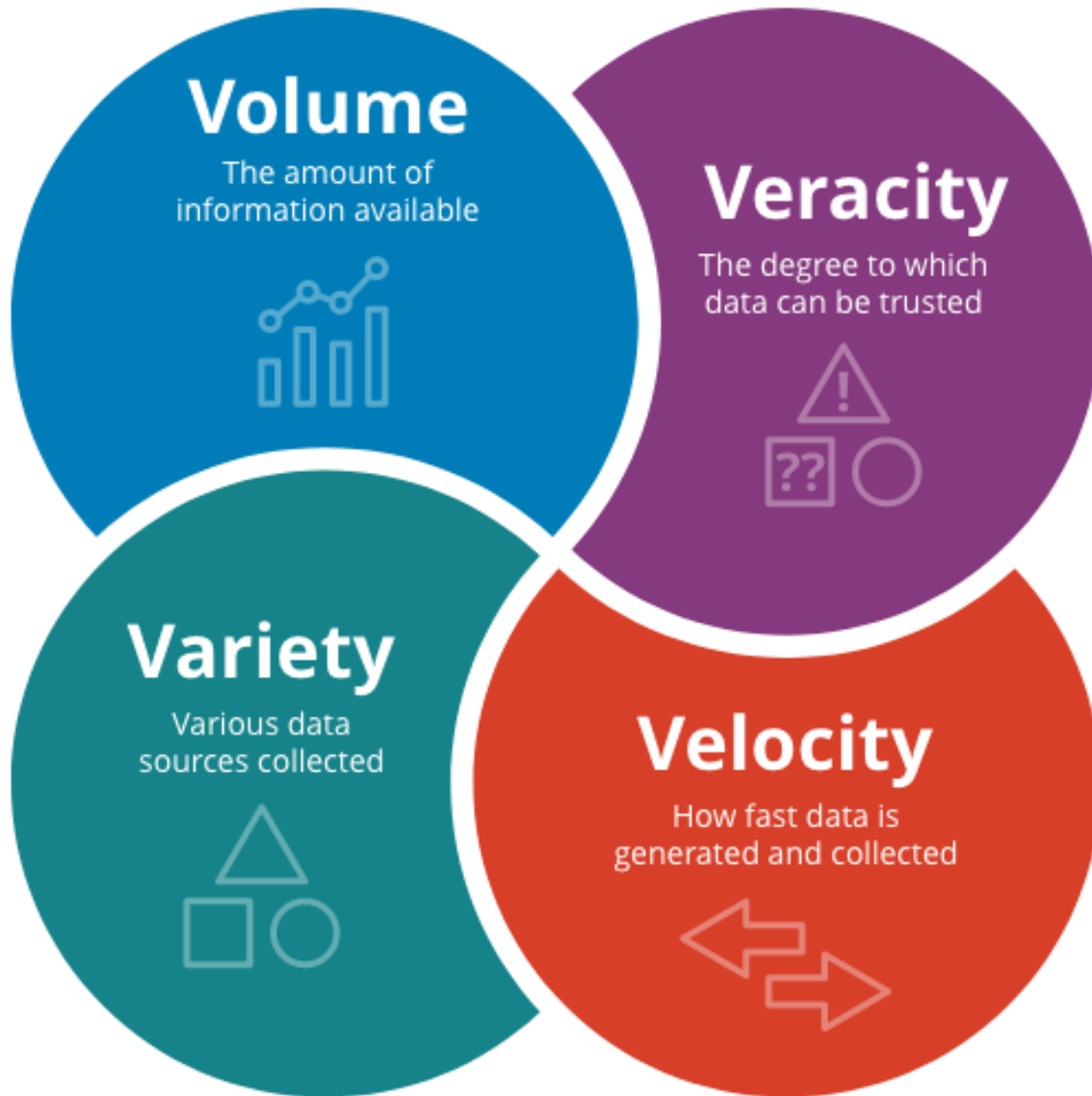
Recolectarlos **Compartirlos**

Buscarlos **DATOS** **Analizarlos**

Visualizarlos **Procesarlos**

Entenderlos

4 V's Of Big Data



A pesar de que puede haber cierta controversia en cuanto a qué puede considerarse Big Data y qué no, prácticamente todos los analistas profesionales y especialistas coinciden en que el Big Data tiene cuatro dimensiones con unas características propias: volumen, variedad, velocidad y veracidad.

Las 4 V de Big Data

#1 Volumen

Tradicionalmente, los datos se han venido generando de forma manual.

Ahora provienen de máquinas o dispositivos y se gestan de manera automática, por lo que **el volumen a analizar es masivo**. Esta característica del Big Data se refiere al tamaño de las cantidades de datos que se generan actualmente.

Las cifras son abrumadoras.

Y es que los datos que se producen en el mundo durante dos días equivalen a todos los generados antes del año 2003.

Estos grandes volúmenes de datos que se producen a cada momento suponen **retos técnicos y analíticos importantes** para las empresas que los gestionan.



Las 4 V de Big Data

#2 Variedad

El origen de los datos es altamente heterogéneo.

Proviene de múltiples soportes, herramientas y plataformas: cámaras, smartphones, coches, sistemas GPS, [redes sociales](#), registros de viajes, movimientos bancarios, etc.

A diferencia de hace unos años, cuando los datos que se almacenaban se extraían, principalmente, de hojas de cálculo y bases de datos.

Los datos que se recopilan pueden venir estructurados (son más fáciles de gestionar) o no estructurados (en forma de documentos, vídeos, mensajes de correo electrónico, redes sociales, etc.).

Dependiendo de esta diferenciación, cada tipo de información se tratará de manera distinta, a través de unas herramientas específicas.

La esencia del Big Data reside en, posteriormente, combinar y configurar unos datos con otros.



Las 4 V de Big Data

#3 Velocidad

El flujo de datos es masivo y constante.

En el entorno del Big Data, los datos se generan y almacenan a una velocidad sin precedentes.

Este gran volumen provoca que los datos queden desfasados rápidamente y que pierdan su valor cuando aparecen otros nuevos.

Las empresas, por lo tanto, deben reaccionar muy rápido para poder recopilarlos, almacenarlos y procesarlos.

El reto para el área de tecnología es almacenar y gestionar grandes cantidades de datos que se generan continuamente.

El resto de áreas también deben trabajar a gran velocidad para convertir esos datos en información útil antes de que pierdan su valor.



Las 4 V de Big Data

#4 Veracidad

Esta característica del Big Data probablemente sea la que supone un mayor reto.

El gran volumen de datos que se genera puede hacer que dudemos del grado de veracidad de todos ellos, ya que la gran variedad de los datos provoca que muchos de ellos lleguen incompletos o incorrectos.

Esto se debe a múltiples factores, por ejemplo, si los datos provienen de distintos países o si los proveedores utilizan diferentes formatos. Estos datos deben ser limpiados y analizados, una actividad incesante ya que continuamente se generan otros nuevos.

La incertidumbre en cuanto a la veracidad de los datos puede provocar ciertas dudas sobre su calidad y su disponibilidad en un futuro.

Por ello, las empresas deben asegurarse de que los datos que están recopilando tengan validez, es decir, que sean los adecuados para los objetivos que se pretenden alcanzar con ellos.



Rompiendo paradigmas: La 5ta V

#5 Valor



Esta característica representa el **aspecto más relevante** de Big Data.

El valor que generan los datos, una vez convertidos en información, puede considerarse el aspecto más importante. Con ese valor, las empresas tienen **la oportunidad de sacar el máximo partido a los datos** para introducir mejoras en su gestión, **definir estrategias más óptimas**, obtener una clara ventaja competitiva, realizar ofertas personalizadas a los clientes, aumentar la relación con el público, y mucho más.

Para ser conscientes de todas las oportunidades que se pueden extraer mediante la aplicación del Big Data es necesario entender cuáles son **los principales elementos que le aportan valor** y que provocan que su aplicación a nivel empresarial sea una apuesta segura.

La Paradoja de los Datos

Solo se analiza el **10%** de los datos que recopilan las organizaciones

Cuanto más datos recopile una empresa, **menos probable es que los analice**



Total Time to Insight

Total Time To Insight (TTTI) es el tiempo que transcurre desde la generación de datos hasta una información útil basada en esos datos.



Roles en Big Data



Data Analyst/Business Analyst

Se especializa en el negocio y en realizar análisis de datos, extrayendo conclusiones.

En general no tienen mucho conocimiento técnico, pero utilizan herramientas de explotación y análisis gráficos para realizar su trabajo.

Habilidades: los analistas de datos deben tener una comprensión básica de algunas habilidades: recopilación de datos, visualización de datos, análisis exploratorio de datos.

Herramientas: Microstrategy, Qlik, Tableau, Power BI entre otras.

Roles en Big Data

Data Architect - orientado a modelado de datos (Data Modeler)

Se especializa en el modelado de datos, tanto relacional o no relacional y fundamentalmente dimensional.

Definen cómo los datos serán ingestados y almacenados, conformados e integrados, administrados y explotados por diferentes aplicaciones.

Responsable de la capa semántica para presentar la información en términos de negocio.

Herramientas: MicroStrategy, Governance Catalog, Bus Matrix (para documentar hechos y dimensiones), Diagrama ER en Erwin



Data Architect

Roles en Big Data

Data Architect - orientado a tecnologías y plataformas

Se especializa en el conocimiento de tecnologías desarrolladas para Big Data, encargándose de diseñar la topología, definir las soluciones, configurar, mantener y soportar las plataformas tecnológicas.

Definen cómo los datos serán ingestados, almacenados, consumidos, integrados, administrados y explotados por diferentes aplicaciones, asegurando la integración de todas las plataformas con las bases de datos y tecnologías tradicionales existentes en la organización.

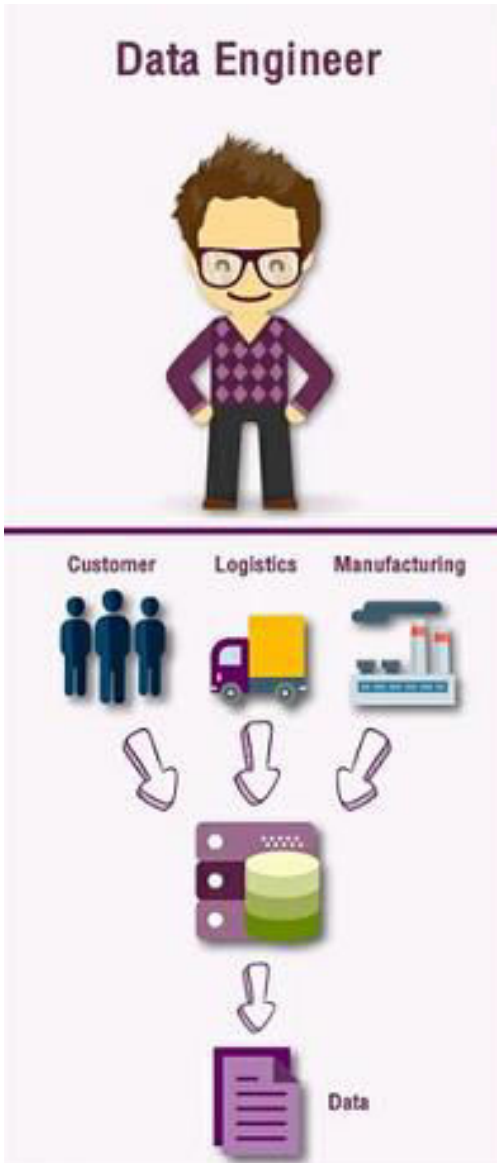
Habilidades: Tiene conocimiento avanzado en plataformas de almacenamiento y procesamiento de datos distribuidos tanto on premise como cloud. Exhaustivo conocimiento de la arquitectura de la base de datos NoSQL. Conocimiento de Herramientas de Extracción, transformación y carga (ETL/ELT) Batch, microbatch, near real time y real time.

Herramientas: Plataforma Hadoop, Nifi, Kafka, Spark, MapReduce, Hive, MongoDB, otras Bases de datos relacionales y nosql, entre otros



DATA ARCHITECT

Roles en Big Data



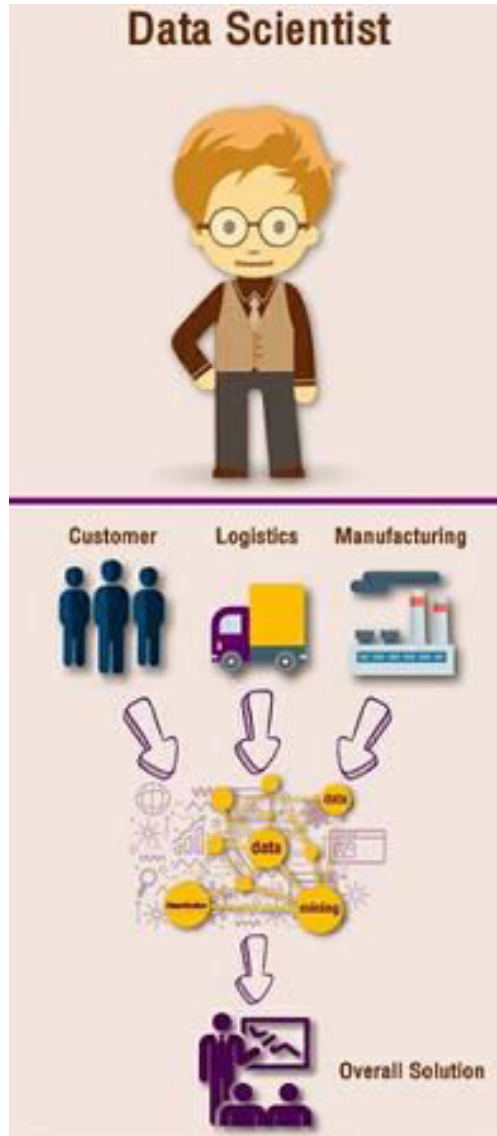
Se encargan de analizar requerimientos y necesidades, investigar potenciales fuentes de datos, diseñar modelos y procesos de carga/ingesta. Entre sus principales tareas se incluye investigar fuentes de datos, descubrir patrones y relaciones, modelar dimensionalmente procesos de negocio, establecer los mapeos desde las fuentes de datos hasta el modelo propuesto, diseñar procesos de preparación de datos contemplando la extracción, limpieza, conformación y carga para su posterior uso analítico u operacional.

Entre sus **principales tareas** se incluye la **preparación de datos para su posterior uso analítico u operacional**. Esta preparación implica la de construir y automatizar pipelines para buscar datos de diferentes fuentes, ingestarlos, integrarlos, consolidarlos, limpiarlos y enriquecerlos para su uso posterior.

Habilidades: Conocimiento de sistemas de base de Datos Relacionales y noSQL. Modelado de datos y herramientas ETL/ELT Batch y Near real time. Desarrollo de APIs para acceso a datos. Desarrollar y utilizar plataformas de almacenamiento y procesamiento distribuido.

Herramientas: BD Relacionales, BD NoSQL, Hadoop, Spark, MapReduce, SQL, HivE, Sqoop, Kafka, Nifi, entre otros. Lenguajes: Python, JavaScript, Java, Scala entre otros

Roles en Big Data



Se encarga de realizar análisis y extracción de conocimiento y conclusiones a partir de grandes volúmenes de datos. Para realizar dicho análisis se basa en utilizar y combinar la informática, con matemáticas y estadística, debiendo comunicar el conocimiento extraído de los datos y la visión de negocio.

Habilidades: Conocimiento de Procesamiento distribuido, algoritmos predictivos, matemáticas, estadísticas, algoritmos orientados a machine learning, buena comunicación y conocimiento del negocio, entre otros.

Herramientas: BD Relacionales, BD NoSQL, Hadoop, Spark, MapReduce, Hive, Notebook Jupyter, Notebook Zeppelin, entre otros. Lenguajes: R, Python, Scala.

Roles en Big Data

Data Science Roles & How They Interact

Enable data access & utilization & enable value capture

Builds and supports the infrastructure or 'data pipe' and all associated SW engineering infrastructure tasks.

Data Engineer

Core skills:



Collect, manage, analyze & visualize data

Build, deploy data infrastructure & architecture

Journey Maps

Value Streams & Service Maps

Deployment & Integration

Core skills:



Help the business make better decisions through data

Blend of business, analytic and math skills to explore and solve challenges, bridging the data and business communities.

Business Stakeholder

Data Scientist

Core skills:



Analysis
Programming
Big Data

Ascribes value to raw data through original interpretation & modeling

Use sophisticated methods to interrogate the data

Personas

Storyboards

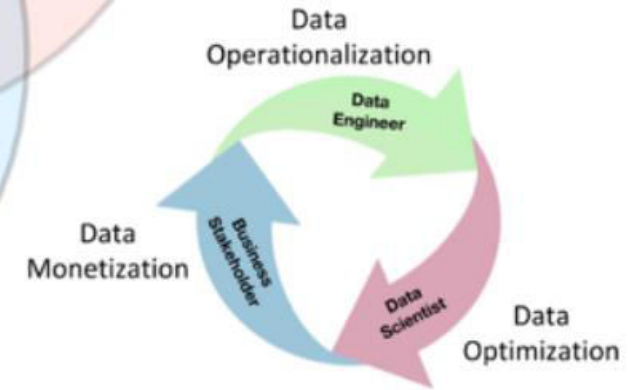
Variables & Metrics to Test

Optimize & enable data for business & functional value capture & value creation

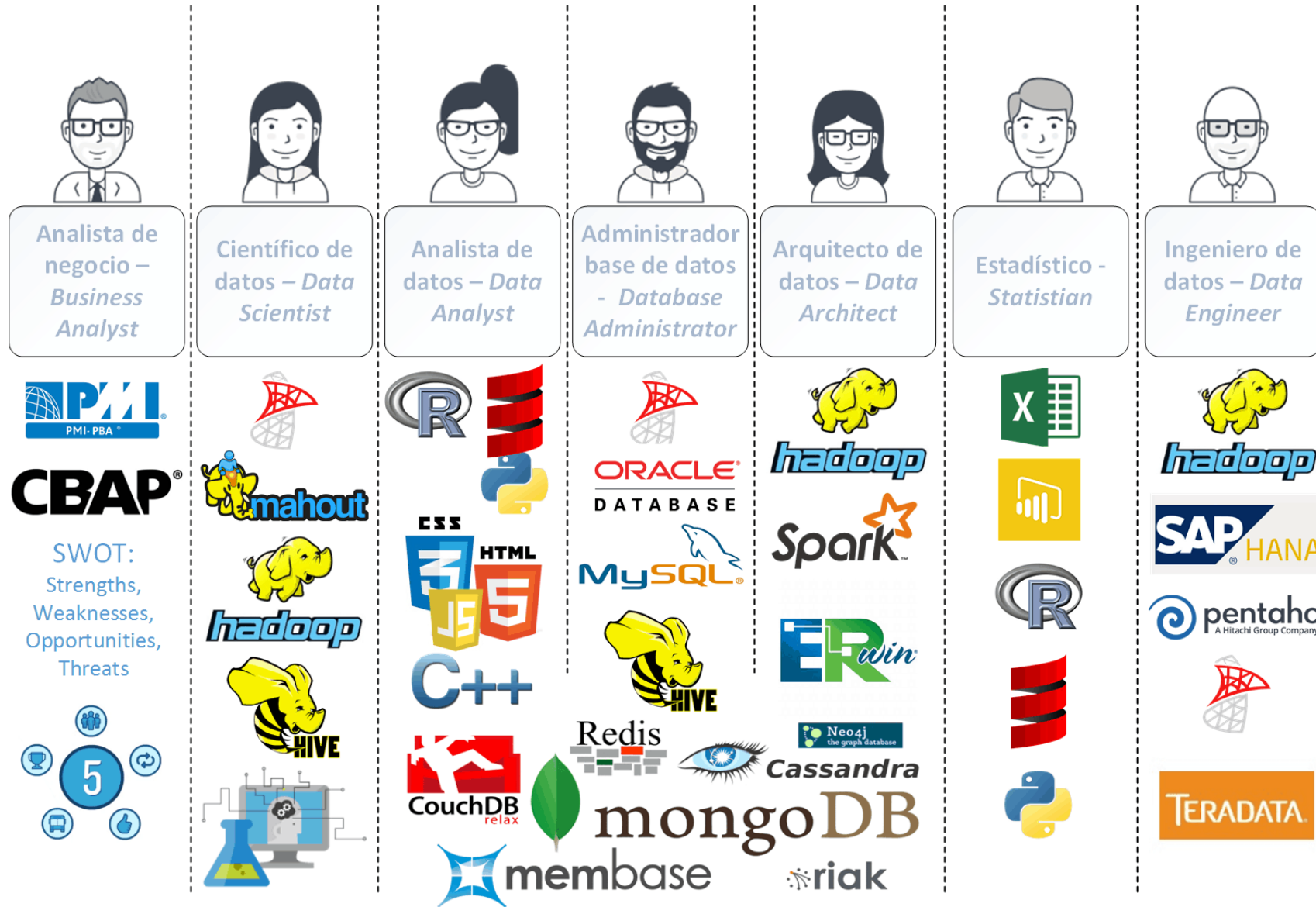
Analysis and interpretation of complex digital data to extract or discover knowledge and assist decision-making.



Hypothesis Development
Monetization
Governance



Roles en Big Data



Roles en Big Data



TP N° 1: Ejercicio de búsqueda: casos posibles de uso de Big Data



- Identifique al menos 3 casos posibles de uso de Big Data. Pueden ser sobre procesos/actividades cercanos a su espacio de trabajo y/o estudio o buscar casos reales en internet.
- Para cada caso desarrolle brevemente cómo lo usaría (cuál sería el beneficio específico del uso de las tecnologías del Big Data)
- Especifique cuáles serían los orígenes de los datos que necesitaría para su caso (por ejemplo CRM, Redes Sociales, IoT, Información de Geoposicionamiento, etc.)

Armar una presentación en 1 hoja en base a lo solicitado. Éxitos!!

TP N° 2: Bases de Datos Relaciones vs No Relacionales

- Detallar 3 características diferenciales de las bases de datos relacionales y no relacionales (también conocidas como SQL y NoSQL).
- Buscar ejemplos de cada una de ellas
- Detallar la estructura de las principales base de datos no relacionales

Armar una presentación en 1 hoja en base a lo solicitado. Éxitos!!



BIG DATA & ANALYTICS I



TEMARIO

Módulo 1: Surgimiento de Big Data

Módulo 2: Los Datos Nos Rodean

**Módulo 3: Obtención de los Datos y
Proceso de Toma de Decisión**

Módulo 4: Visualizar con Datos

Disertantes: Lic. Maria Trinidad Aquino – Ing. Raúl Alejandro Grassi

CePETel

Sindicato de los Profesionales
de las Telecomunicaciones

SECRETARÍA TÉCNICA



Instituto Profesional de
Estudios e Investigación

BIG DATA & ANALYTICS I

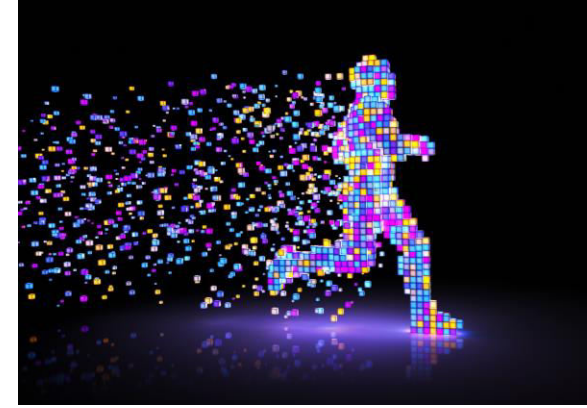
Material Adicional 1: La Ética del Data

La Ética del Data

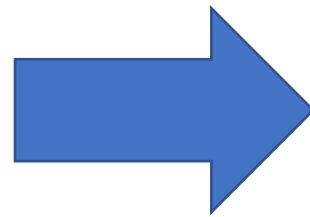


- **A medida que aumenta el uso y el potencial del Big Data, crece el debate sobre la ética de los datos. La tecnología no es el problema, sino cómo la instruimos y para qué la utilizamos.**
- Imagine que **todo lo que hace cada día está monitorizado**. Qué compra en Internet. Qué y con quién habla en las redes sociales. Qué fotos ha hecho. Qué vídeos o series ve. En qué sitio está en ese momento (geolocalización móvil y reconocimiento facial). Qué libros lee. Qué facturas tiene y si es buen pagador o no... Todo. Y ahora imagine que toda esa información conjunta se utiliza **para asignar un rating a cada ciudadano**. Una calificación que influirá en su posibilidad de obtener un crédito con ciertas condiciones y, quizá en el futuro, también en su posibilidad de conseguir o no ciertos empleos, ir a determinada universidad o acceder a una vivienda concreta. ¿Ciencia ficción? No.

La Ética del Data



La ética debe estar presente en todo el ciclo de vida del dato: captación, gestión, privacidad y uso



No hay un estándar sobre lo que los consumidores consideran un uso aceptable de los datos; varía por culturas y tipo de información

La Ética del Data



Las organizaciones que actúen con ética y transparencia en el uso del Big Data serán bien vistas por la sociedad



La Ética del Data Inteligencia Artificial



¿Cómo hacemos buenas a las máquinas?

El debate sobre la ética del dato se acrecienta a medida que avanzamos del Big Data al siguiente estadio tecnológico, la Inteligencia Artificial. En realidad, ambas vienen de la mano, porque no se entiende Inteligencia Artificial sin Big Data, que es la que alimenta a la primera. Pensemos por ejemplo en el coche autónomo, un claro ejemplo de Inteligencia Artificial que no podría funcionar si el vehículo no recibiera miles de millones de datos en tiempo real para “ver y analizar” todo lo que sucede en la carretera y tomar decisiones al respecto.

En lo relativo a la Inteligencia Artificial, el debate en torno a la ética es mucho más amplio y profundo. Porque hoy las decisiones las toman las personas. Pero en el futuro serán las máquinas, que actuarán teniendo en cuenta los datos, los algoritmos programados y su propio aprendizaje (Machine Learning). Y para que tomen buenas decisiones, tendrán que estar bien instruidas. De ahí que algunos expertos del MIT sugieran hablar de Machine Teaching en lugar de Machine Learning poniendo así el foco no en las máquinas, sino en las personas que les enseñan.

La Ética del Data

Inteligencia Artificial



De momento, el debate sobre la ética en la Inteligencia Artificial **no se ha traducido en leyes**, pese a que varios países como China, Estados Unidos, Reino Unido y Emiratos Árabes se han propuesto liderar este campo en el horizonte de 2030. [Emiratos incluso tiene un ministro de Inteligencia Artificial.](#)

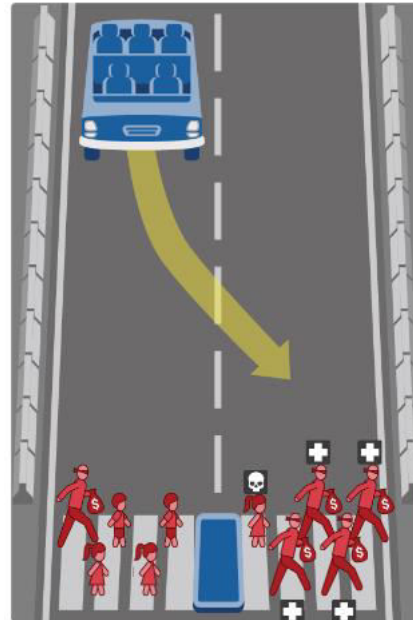
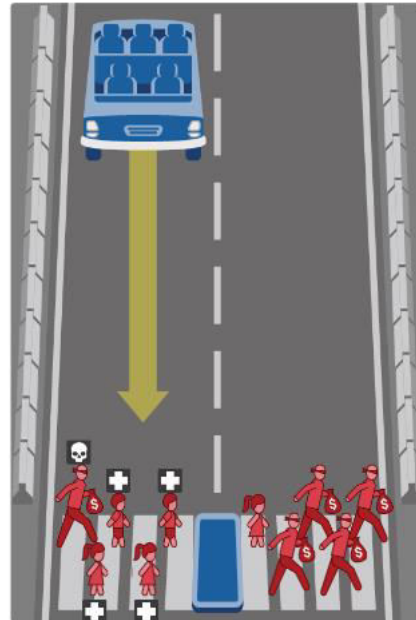
Sí que existen algunos **códigos de buenas prácticas a nivel internacional** e incluso de **ciertas organizaciones**. Pero de momento, el debate sigue siendo **liderado por prestigiosas universidades y organizaciones** de todo el mundo. Varios ejemplos interesantes: [Ethics and Governance of Artificial Intelligence Fund](#), una asociación pilotada por la Universidad de Harvard y el MIT; [Partnership on AI to benefit people and society](#), fundada por Amazon, Apple, Facebook, Google, IBM y Microsoft; [OpenIA](#), una organización sin ánimo de lucro en la que participan compañías como Amazon, Microsoft e Infosys. La universidad de Stanford, con la participación de expertos, está liderando el proyecto [One hundred year study on artificial intelligence](#), con el compromiso de estudiar y anticipar durante cien años los efectos de las tecnologías cognitivas en la sociedad y con [este primer informe](#) como resultado.

La Ética del Data Inteligencia Artificial



¿Qué haría usted si condujera un coche autónomo y tuviera que enfrentarse a estas delicadas situaciones?

Link: [Test Moral Machine](#)



BIG DATA & ANALYTICS I



TEMARIO

- Módulo 1: Surgimiento de Big Data**
- Módulo 2: Los Datos Nos Rodean**
- Módulo 3: Obtención de los Datos y Proceso de Toma de Decisión**
- Módulo 4: Visualizar con Datos**

Disertantes: Lic. Maria Trinidad Aquino – Ing. Raúl Alejandro Grassi

CePETel

Sindicato de los Profesionales
de las Telecomunicaciones

SECRETARÍA TÉCNICA



Instituto Profesional de
Estudios e Investigación

BIG DATA & ANALYTICS I

Módulo 2: Los Datos Nos Rodean

Implementando BIG DATA

Las **tres características principales** de los datos que hacen necesario aplicar técnicas de Big Data en un proyecto de análisis de datos son:

- **Grandes volúmenes de datos**
- **la necesidad de procesar esos datos de manera rápida, y**
- **el uso de tipos de datos complejos.**

En cualquier caso, eso se traduce en la necesidad de una **gran capacidad computacional**, para así poder satisfacer los requerimientos analíticos de los usuarios.

En una solución de **Business Intelligence (BI)** tradicional, se trabaja en la gran mayoría de casos con una arquitectura basada en un único servidor. Este servidor, a pesar de poder ser multi-core, no puede paralelizar una tarea a alto nivel, con lo que debe asumir toda la carga computacional. El resultado es la ejecución de largos procesos para resolver cálculos analíticos.

Big Data se basa en el paradigma de la computación distribuida. Es decir, existe una red computacional (llamada cluster), formada por un conjunto de ordenadores (llamados nodos), que trabajan de manera organizada y colaborativa para resolver el mismo problema.

Implementando BIG DATA

Condiciones fundamentales

ESCALAMIENTO HORIZONTAL



ALTA DISPONIBILIDAD



DISTRIBUCIÓN DE DATOS



PROCESAMIENTO DISTRIBUIDO

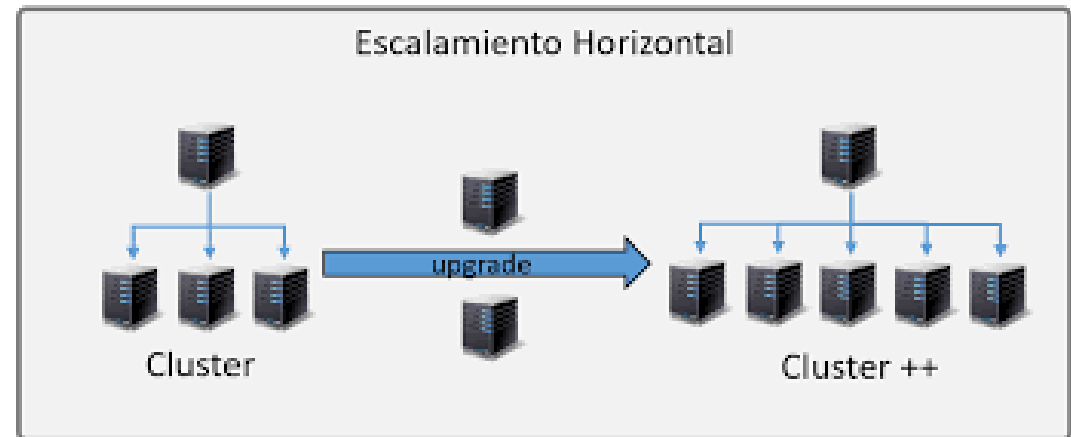


Implementando BIG DATA

Escalamiento Horizontal

Cuando un sistema informático puede llegar a procesar **grandes cantidades de datos**, es muy importante que éste **sea escalable**. Es decir, **que pueda crecer para soportar cargas de trabajo mayores a las que originalmente soporta**.

Una solución Big Data es un sistema escalable, puesto que **un clúster puede aumentar el número de nodos existentes**. Sin embargo, no todo sistema escalable requiere de una solución Big Data.



Implementando BIG DATA

Alta Disponibilidad



Un **clúster de alta disponibilidad** es un conjunto de dos o más máquinas que se caracterizan por mantener una serie de servicios compartidos y por estar constantemente monitorizándose entre sí.

La **redundancia** es la capacidad de interconectar los Data Centers en caso de que el principal falle y no perder tiempo de actividad; mientras que la **alta disponibilidad** es la capacidad de tener operando tus sistemas sin interrupción

La configuración de los sistemas HA consta de dos **componentes**: el **principal** (que es que esta dando servicio continuo) y **secundario** (que es un **clon** del anterior para que de el mismo servicio) **tiene que ser totalmente idénticos** para que el usuario final no se percate de la caída del sistema y pueda seguir usándolo de manera transparente.

Implementando BIG DATA

Distribución de Datos

Se define como sistema distribuido a un **conjunto o grupo de equipos** que son independientes entre sí y que **actúan como un único equipo** de forma transparente y que tienen como objetivo la descentralización del procesamiento o el almacenamiento de información.



Implementando BIG DATA

Procesamiento Distribuido

Cuando **una tarea** llega al cluster de Big Data, ésta **se descompone en un conjunto de subtareas**, que son asignadas a los **distintos nodos del cluster**.

Los **nodos trabajan** entonces en **paralelo** (a la vez) en cada una de sus tareas, con lo que **el tiempo de cálculo** del conjunto de tareas **se ve reducido** considerablemente.

Finalmente, **es necesario** un post-proceso para **consolidar los datos antes de ofrecer el resultado a los usuarios**.

El hecho de disponer de una **red de nodos trabajando en paralelo** permite **acelerar el tiempo de proceso total de la tarea** hasta el punto que, a pesar de tener que realizar un trabajo adicional, **la computación distribuida es más eficiente en procesos de cálculos repetitivos y con muchos datos**.



En vez de que la computación se realice sólo en una computadora central, se distribuye mediante redes a los sitios donde se realiza el trabajo de la empresa.

Implementando BIG DATA

Conclusión

La **distribución de la carga de trabajo** entre los **distintos nodos** de un **cluster en una solución Big Data**, permite **realizar las tareas en paralelo**, consiguiendo **grandes mejoras** en la eficiencia de los procesos.

La computación distribuida es más eficiente en escenarios con unos cálculos a realizar sobre un gran conjunto de datos. De esta manera se paraleliza el trabajo y se reduce el tiempo total de cálculo.

A la **hora de diseñar la arquitectura de un sistema Big Data**, es conveniente **tener en cuenta el grado de paralelización de las tareas**, que será función del número de procesadores y cores disponibles en el sistema.

Si el grado de paralelización es bajo, el rendimiento en según qué tareas, puede acabar siendo también muy bajo.



- más eficiente y a un menor coste
- mayor tolerancia a los fallos
- mayor rendimiento
- mayor velocidad en el procesamiento

Fuentes de Datos

Fuentes Internas:

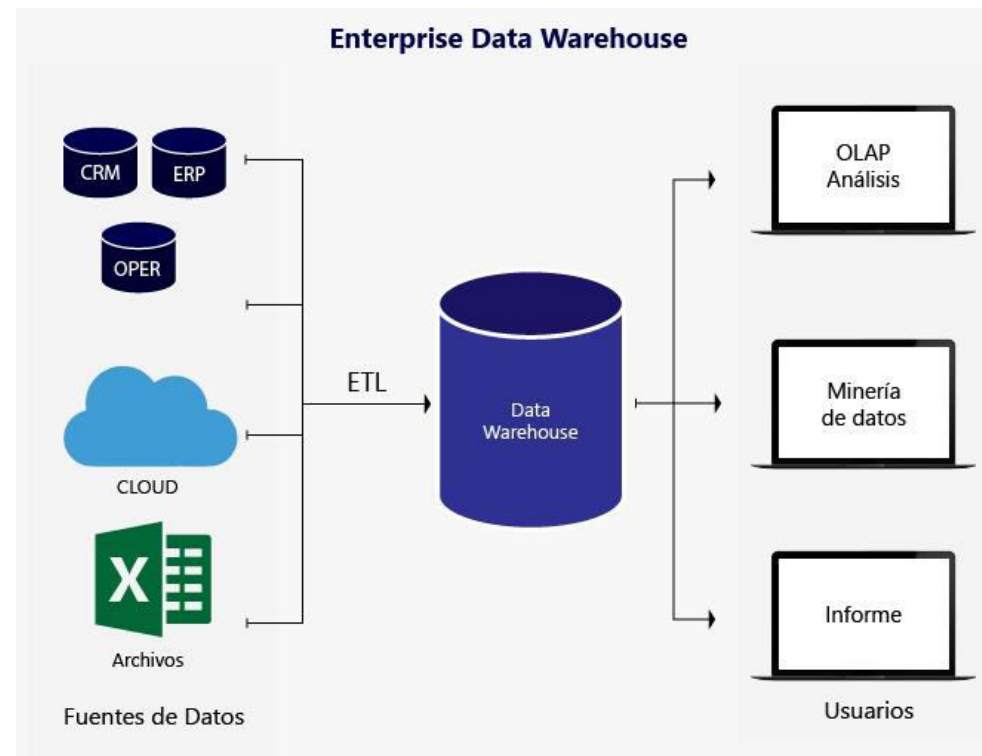
- ERP
- CRM
- SCM
- MES
- BPM
- Sistemas Legacy
- Planillas de Cálculo –
- Stream Events (Sensores, medidores)

Formatos

- Bases de Datos Relacionales
- Bases de Datos NoSQL
- Archivos (CSV, AVRO, JSON, XML, XLS, VSAM, comprimidos -zip, zlib, bzip-)

Fuentes Externas a la Organización

- Redes Sociales
- Datos de Cámaras empresariales
- Open Data - Gobierno
- Web Crawling Data



SQL vs. NoSQL

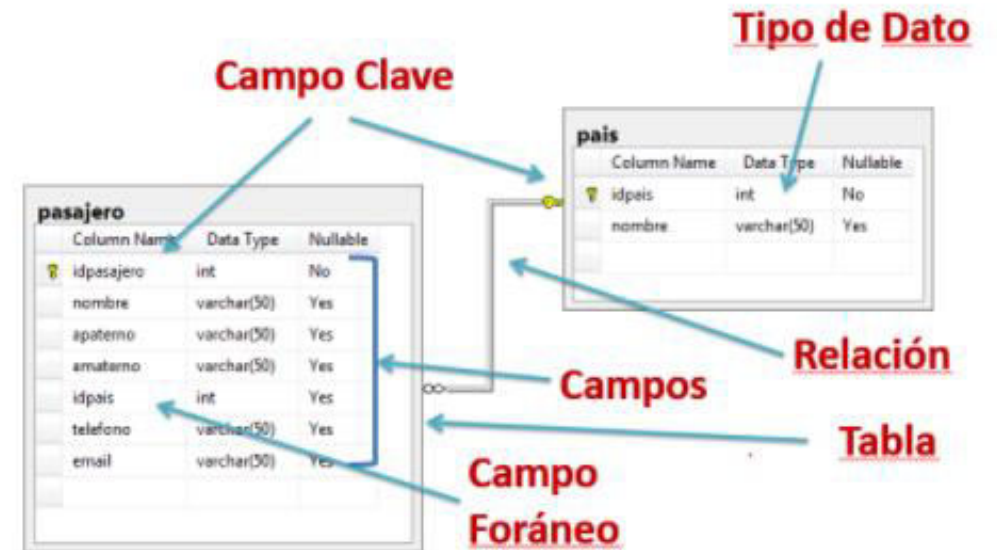
Bases de datos relacional o SQL

¿Qué es un sistema de base de datos relacional?

El modelo relacional es una forma intuitiva y directa de representar datos. Las bases de datos relacionales son el modelo más utilizado actualmente

Una base de datos relacional es, en esencia, un conjunto de tablas (o relaciones) formadas por filas (registros) y columnas (campos); así, cada registro (cada fila) tiene una ID única, denominada clave y las columnas de la tabla contienen los atributos de los datos.

Una de las principales características de la base de datos relacional es evitar la duplicidad de registros y a su vez garantizar la integridad referencial, es decir, que si se elimina uno de los registros, la integridad de los registros restantes no será afectada. Además, gracias a las claves se puede acceder de forma sencilla a la información y recuperarla en cualquier momento.



Implementando BIG DATA

BASES DE DATOS NoSQL

¿Qué es NoSQL?

Sistemas de gestión de bases de datos que difieren del modelo clásico de bases de datos relacionales: no usan SQL como lenguaje de consulta, los datos almacenados no requieren estructuras fijas como tablas, no garantizan consistencia plena y escalan horizontalmente.

Not Only SQL

Surgieron para complementar a las bases de datos tradicionales, no para reemplazarlas.



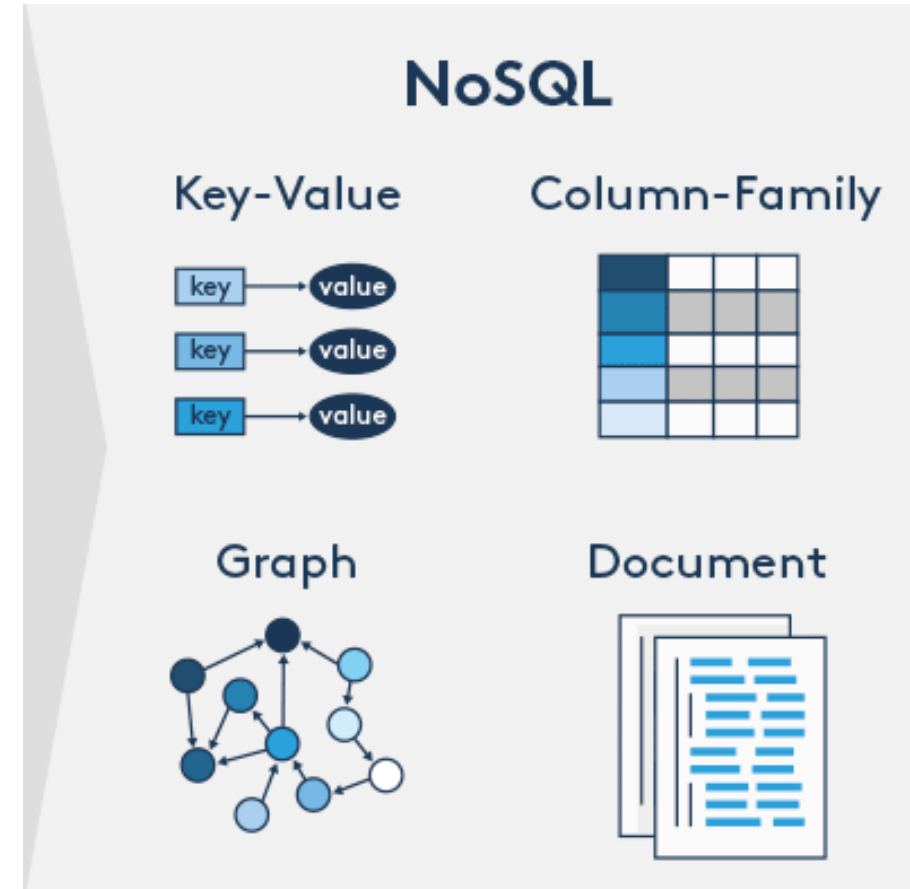
Implementando BIG DATA

BASES DE DATOS NoSQL

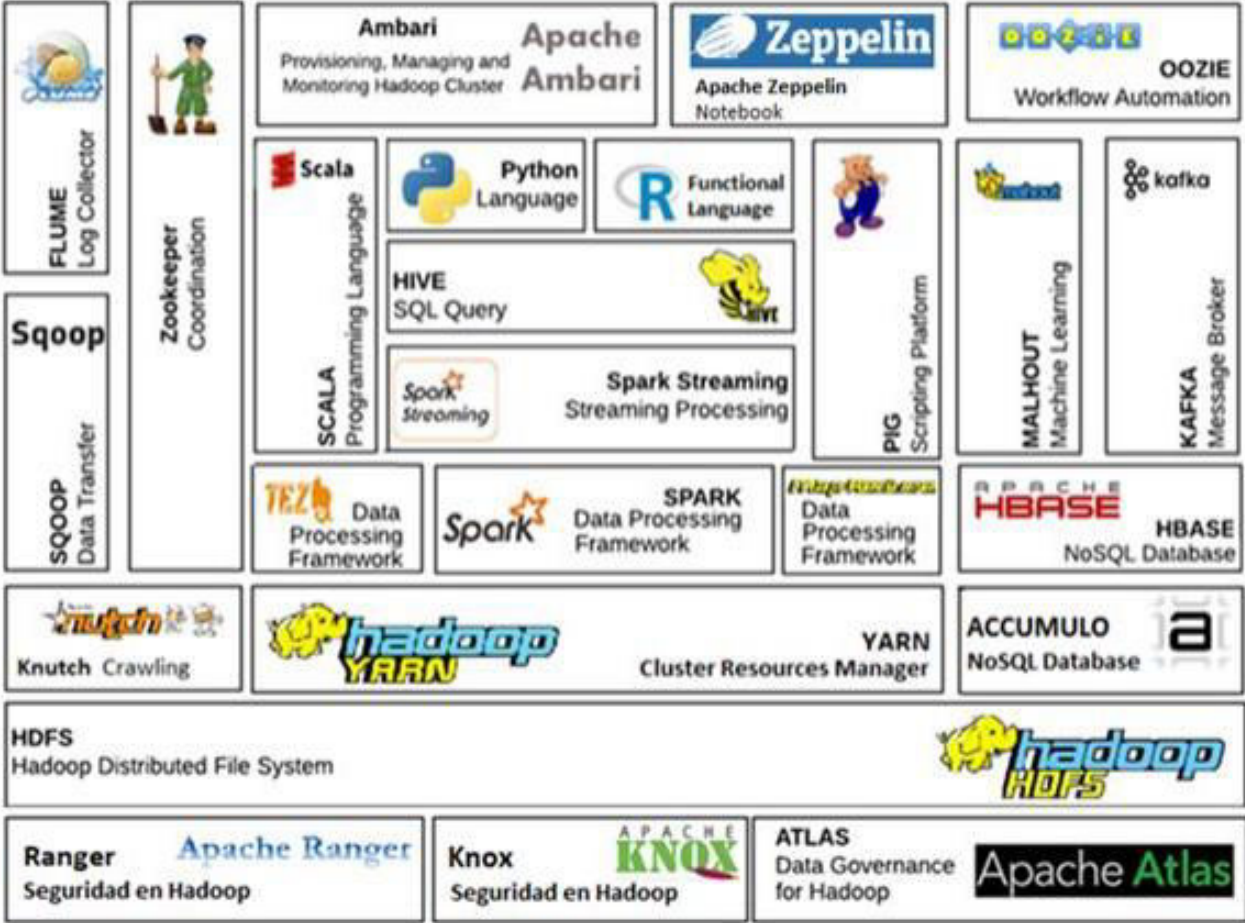
Tipos de Bases NoSQL

Generalmente, las bases de datos NoSQL se clasifican según su forma de almacenar los datos, y comprenden categorías como:

- Bases de datos clave-valor
- Bases de datos columnares
- Bases de datos documentales
- Bases de datos orientadas a grafos.



Hadoop - Definición

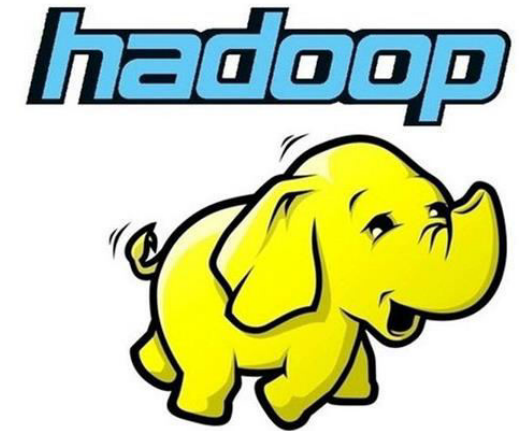


¿Qué es Hadoop?

Hadoop es un sistema de **código abierto** que se utiliza para almacenar, procesar en paralelo y analizar grandes volúmenes de datos.

Hadoop - Principales Características

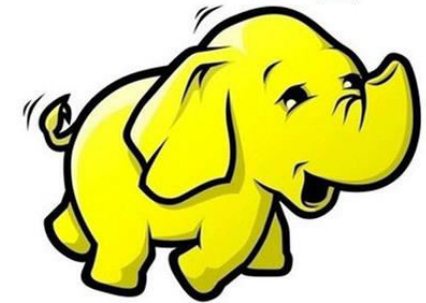
- **Aísla** a los desarrolladores de todas las **dificultades** presentes en la programación paralela.
- Cuenta con un ecosistema que sirve de gran ayuda al usuario, ya que permite **distribuir los archivos en nodos**, que no son otra cosa que servidores con commodity-hardware.
- Es capaz de **ejecutar procesos en paralelo** en todo momento.



Hadoop - Principales Características

- Dispone de **módulos de control** para la **monitorización de los datos**.
- Cuenta con **distintas herramientas** para realizar **consultas de los datos**.
- También **potencia la aparición de distintos add-ons**, que facilitan el trabajo, **manipulación y seguimiento** de toda la **información** que en él se **almacena**.

hadoop



Hadoop - Cluster

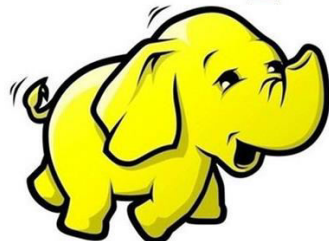
La configuración habitual de Hadoop es tenerlo en un **cluste de máquinas**.

De manera que debemos tener una máquina “maestra” y “n” cantidad de máquinas “esclavas”.

La idea principal es que la máquina maestra **gestionará** toda las **tareas** y las **enviará** a las máquinas “esclavas”.

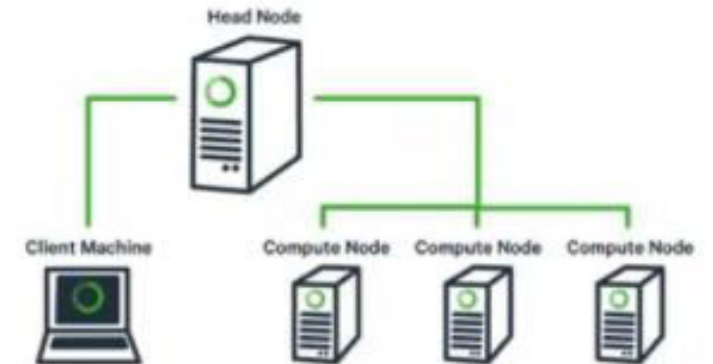
Esas máquinas realizarán todo el **procesamiento de los datos** y a continuación volverían a **informar** a la máquina “maestra”.

hadoop

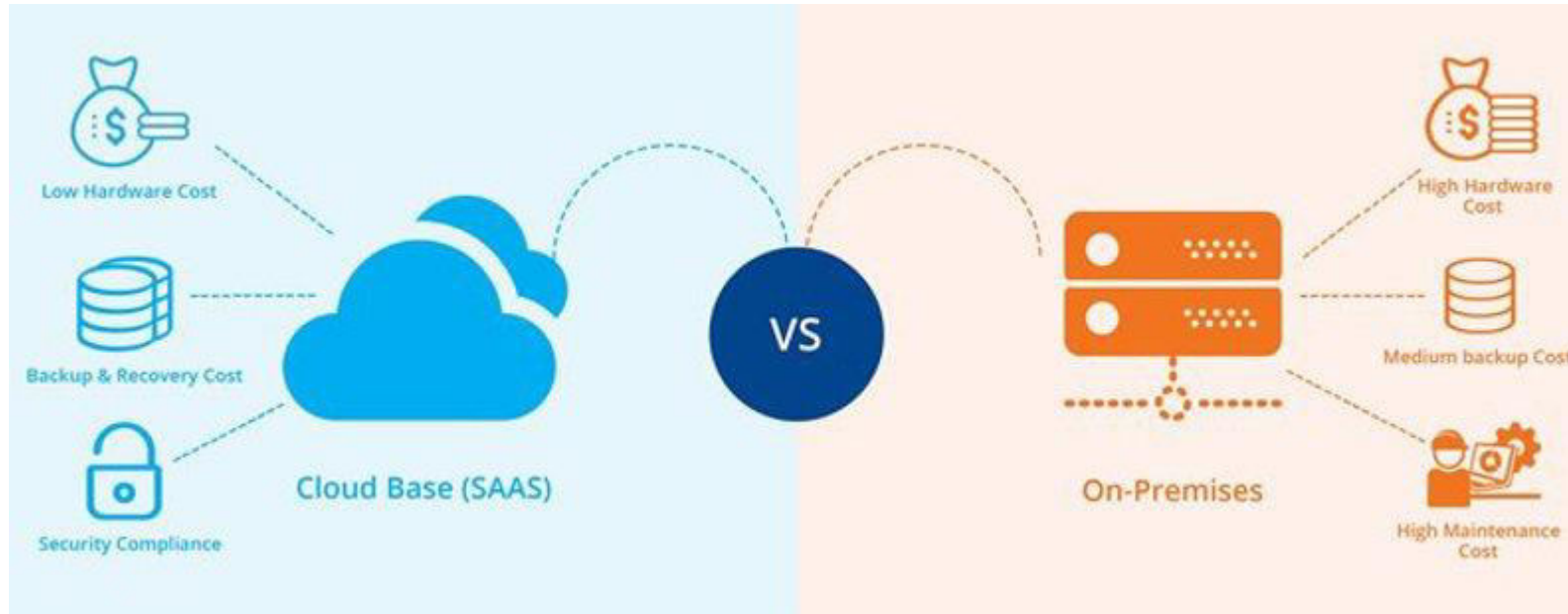


Cluster:

- ▶ 1 Máquina maestra.
- ▶ N máquinas esclavas.



Cloud vs On Premise



Los programas se instalan y ejecutan sobre los datos en servidores o granjas de servidores remotos en distintos proveedores de servicios web. Por ej. AWS, Microsoft Azure, Google

Los programas se instalan y ejecutan sobre los datos en servidores locales en las instalaciones de la empresa.

[Video: Entendiendo de que se trata On Premise y Cloud Storage](#)

Cloud Services

¿Qué son los servicios en la nube?

Los **servicios en la nube** son servicios que **se utilizan** a través de **Internet**. Es decir, **no están físicamente instalados en tu ordenador**. Se trata de un nuevo paradigma que surgió con el advenimiento de la World Wide Web.

Antes de que apareciera la nube, **todos los programas informáticos se instalaban en el ordenador**. Los servicios en la nube son programas que se alojan en un servidor accesibles desde cualquier dispositivo conectado a Internet.

Las **ventajas** de este tipo de servicios son evidentes, ya que su **uso no está restringido a un solo equipo informático** y la seguridad, capacidad de almacenamiento y recursos de la nube son mayores que los de un ordenador.



Cloud Services

Tipos de servicios en la nube

Podemos agrupar los servicios de la nube en varias categorías:

•Software as a Service (SaaS)

El más utilizado. El software está alojado en servidores de los proveedores y el cliente accede a ellos a través del navegador web. Todo lo relacionado con mantenimiento, soporte y disponibilidad es manejado por el proveedor.

•Platform as a Service (PaaS)

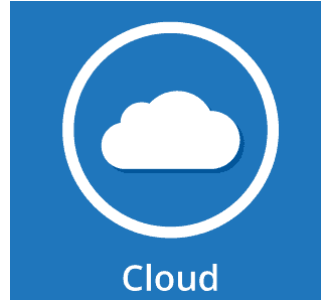
En este tipo de servicios en la nube el proveedor ofrece acceso a un entorno basado en cloud en el cual los usuarios pueden crear y distribuir sus propias aplicaciones. El proveedor proporciona la infraestructura subyacente.

•Infrastructure as a Service (IaaS)

Un proveedor de servicios proporciona el software y las aplicaciones a través de Internet. Los usuarios se suscriben al software y acceden a él a través de la web o las APIs del proveedor.

Un usuario estándar normalmente utilizará solo SAAS.

Cloud



Ventajas	Desventajas
Requisitos de hardware. La PC no requiere un alto rendimiento, ya que se realiza en el centro de datos.	Acceso a Internet. Si no hay conexión a Internet o está inestable, no se puede usar el software.
Mantenimiento. Las actualizaciones las lleva a cabo personal especializado, por lo que no debes preocuparte por el mantenimiento.	Protección de datos. El proveedor es quien maneja tus datos.
Escalabilidad. Se pueden añadir o eliminar funciones o cuentas con relativa rapidez.	Independencia. Si el proveedor deja de operar en la nube, no podrás seguir usando el software.
Costos. Ya que SaaS (software as a Service) es un modelo de suscripción, los costos iniciales son bastante bajos. Pagas por el uso.	

On Premise



Ventajas On premise	Desventajas On premise
Compra. Utilizando un software no se añaden costos.	Hardware. Los dispositivos que se utilizan deben ser adecuados para el software.
Control. Tú eres quien decide qué hacer con el software.	Trabajo adicional. El mantenimiento del capital humano especializado.
Protección de datos. Los datos sensibles pueden permanecer en el sistema y no se traspasan a terceros.	Escalabilidad deficiente. Las licencias son caras y deben adquirirse a largo plazo.
Acceso. Puesto que el software se utiliza en tus instalaciones, siempre puedes trabajar con el programa sin restricciones.	Soporte limitado. El desarrollo del software tras la adquisición suele limitarse en el tiempo.

Resumen comparativo - Cloud vs On Premise



VS



	On Premise	Cloud
Costos	El precio de la licencia es relativamente alto, se paga solo una vez.	El modelo con suscripción supone un pago recurrente, aunque relativamente bajo.
Instalación	Tú debes instalar el software en tu propio hardware.	Puedes acceder al software a través de Internet.
Mantenimiento	Tú eres el encargado de realizar todas las actualizaciones.	Los proveedores del servicio, son los encargados de instalar las actualizaciones.
Escalabilidad	A veces es posible adquirir extensiones para el software, aunque en la mayoría de los casos hay que comprar otro nuevo.	Se pueden añadir o eliminar funciones y usuarios sin esfuerzo.
Hardware	Tú debes contar con el hardware adecuado y asegurarte de que el software es compatible.	Como el software está alojado en servidores especiales, únicamente necesitas disponer de conexión a Internet.
Protección de datos	Los datos permanecen en tu equipo.	El proveedor debe asegurarse de que terceros no autorizados no puedan acceder a tus datos, ni siquiera durante la transmisión en Internet.
Foco	Hadoop se enfoca en manipular grandes sets de datos estructurados y no estructurados	La nube se enfoca en la performance del sistema, seguridad de datos y disponibilidad.

Casos de Uso de Big Data

Conocimiento del cliente & Marketing

- Pasar de un modelo estático a uno dinámico
- Campañas personalizadas.
- Retención de clientes insatisfechos.
- Captar potenciales clientes.
- Mejorar la experiencia de usuario.
- Perfiles para toma de decisiones.
- Anticipar el comportamiento.
- Visión 360 del cliente.



Casos de Uso de Big Data



Publicidad en Tiempo Real / Predicción del Comportamiento

- Establecer patrones de conducta en base a la información que los consumidores vuelcan a la red
- Ofrecer los productos adecuados en el momento indicado, enviando el mensaje adecuado en el momento indicado.

Casos de Uso de Big Data

Omnicanalidad

- Comprar dónde, cuándo y cómo quiera.
- Manejar la misma información en todos sus canales.
- Mantener una personalidad coherente en todos sus canales.
- Tener sus equipos cohesionados
- Mantener una comunicación centralizada y de fácil acceso



Casos de Uso de Big Data

Detección de fraude

- Modelado predictivo según el comportamiento.
- Análisis histórico del cliente y sus denuncias.
- Conductas en redes sociales.
- Lejos de poder ser detectado por el ojo humano

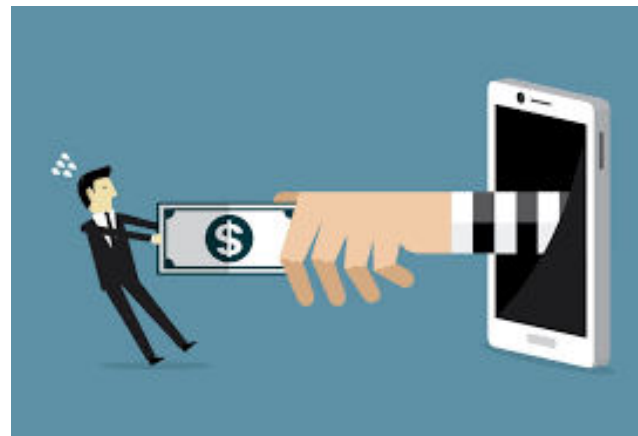


Casos de Uso de Big Data

Detección de fraude



El departamento de legales de Mercado Libre se encontró ante el problema de responsabilizarse frente a los productos que se comercializaban en su portal. Dada esta situación todas las publicaciones debieron ser analizadas para encontrar patrones de fraude o de posible incompatibilidad legal, ya sea venta de artículos ilegales o incluso usuarios que evadían la ley publicando datos personales en sus comentarios o que no vendían realmente el producto anunciado.



Casos de Uso de Big Data

NETFLIX

- Apuesta: De distribuidor a generador de contenido.
- Netflix's Chief Communications Officer:

"Because we have a direct relationship with consumers, we know what people like to watch and that helps us understand how big the interest is going to be for a given show. It gave us some confidence that we could find an audience for a show like House of Cards."

Qué sabían:

- Gran cantidad de usuarios que vieron películas de David Fincher.
- Versión inglesa de "House of Cards" fue muy vista.
- Muchos de esos usuarios también vieron películas de Kevin Spacey y/o David Fincher.

10 trailers distintos.. a medida.

Una producción propia que Netflix creó en base a los gustos de los usuarios, a partir de los patrones de consumo obtenidos de más de 40 millones de consumidores, detectó que lo que más atraía eran contenidos que incluyeran drama, política, sensualidad y poder y, en base a ello, crearon la serie con los actores y directores mas destacados del rubro.



Casos de Uso de Big Data



El uso de herramientas de **IA** no se tiene que limitar a unos pocos casos residuales, sino que puede ser parte clave de la estrategia de una empresa.

La tecnología está ampliamente implementada en la plataforma y llega directamente tanto a los que desean alquilar su espacio como a los que buscan dónde quedarse. El uso de esta rama de la IA abarca “todo, desde cómo recomiendan precios a sus anfitriones, comprendiendo el valor de su espacio, a como hacen el ranking de búsquedas”. Airbnb aplica el **aprendizaje automático** en un amplio rango de campos, entre los que también se encuentra el de la **seguridad** de los que emplean la plataforma, “todas las cosas que hacemos para **detectar fraudes** y mantener a nuestras comunidades a salvo ”.

TP N° 3: Cloud vs. On Premise



- Identificar y describir al menos 3 diferencias entre los servicios Cloud vs On Premise.
- Fundamente las ventajas/desventajas de esas 3 diferencias.

Armar una presentación en 1 hoja en base a lo solicitado para exponer en clase.

Éxitos!!

TP N° 4: Inteligencia Artificial



Diferenciación entre Inteligencia Artificial, Machine Learning y Deep Learning. Ejemplos.

- Detalle las características de una de las tres (a su elección).

Armar una presentación en 1 hoja en base a lo solicitado para exponer en clase. Éxitos!!

BIG DATA & ANALYTICS I



TEMARIO

Módulo 1: Surgimiento de Big Data

Módulo 2: Los Datos Nos Rodean

**Módulo 3: Obtención de los Datos y
Proceso de Toma de Decisión**

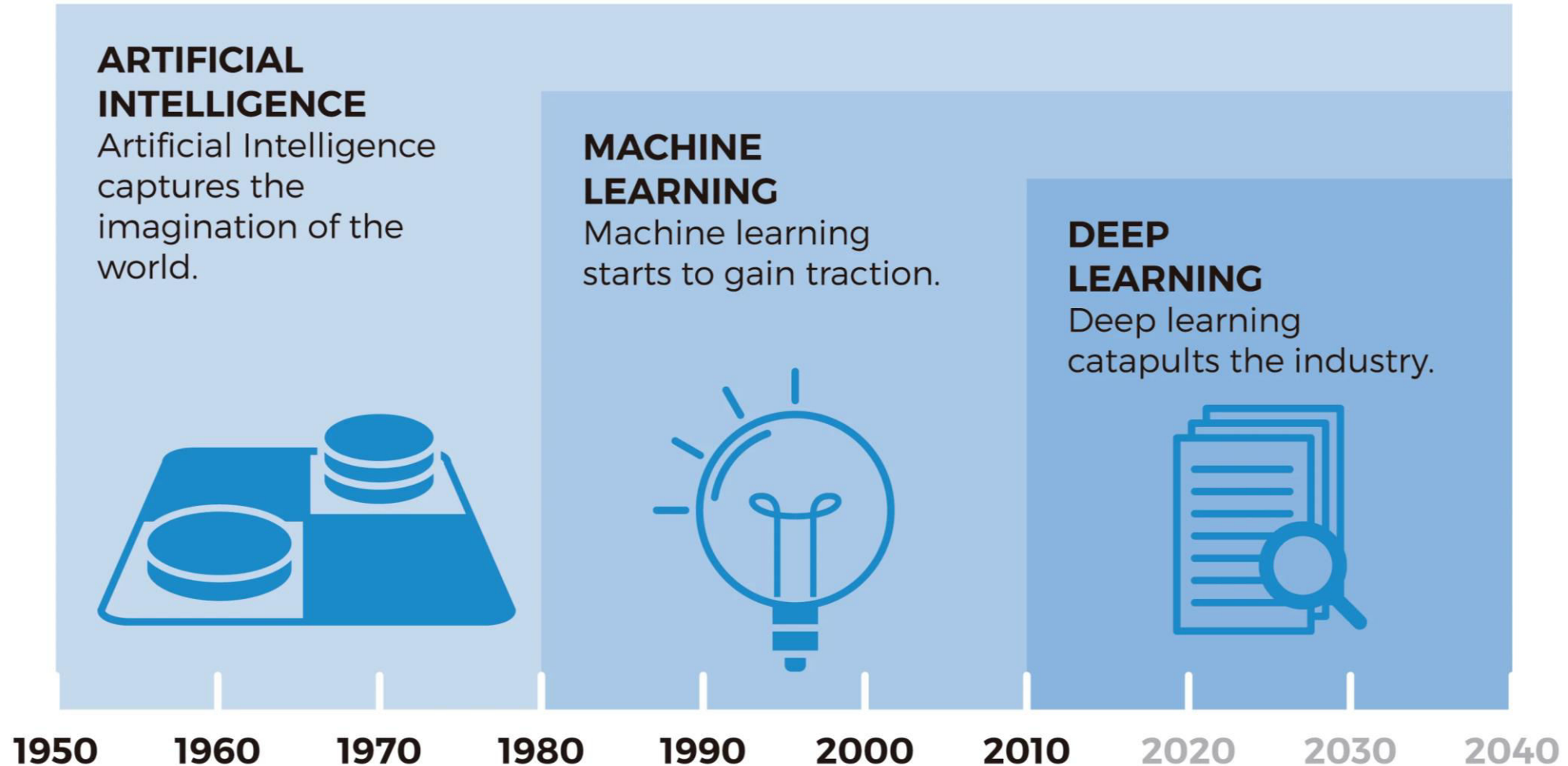
Módulo 4: Visualizar con Datos

Disertantes: Lic. Maria Trinidad Aquino – Ing. Raúl Alejandro Grassi

BIG DATA & ANALYTICS I

Módulo 3: Obtención de los Datos y Proceso de Toma de Decisión

Inteligencia Artificial, Machine Learning y Deep Learning



[Video Introductorio: Mapa Conceptual](#)

Inteligencia Artificial

La **inteligencia artificial** es la habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear.

Inteligencia artificial

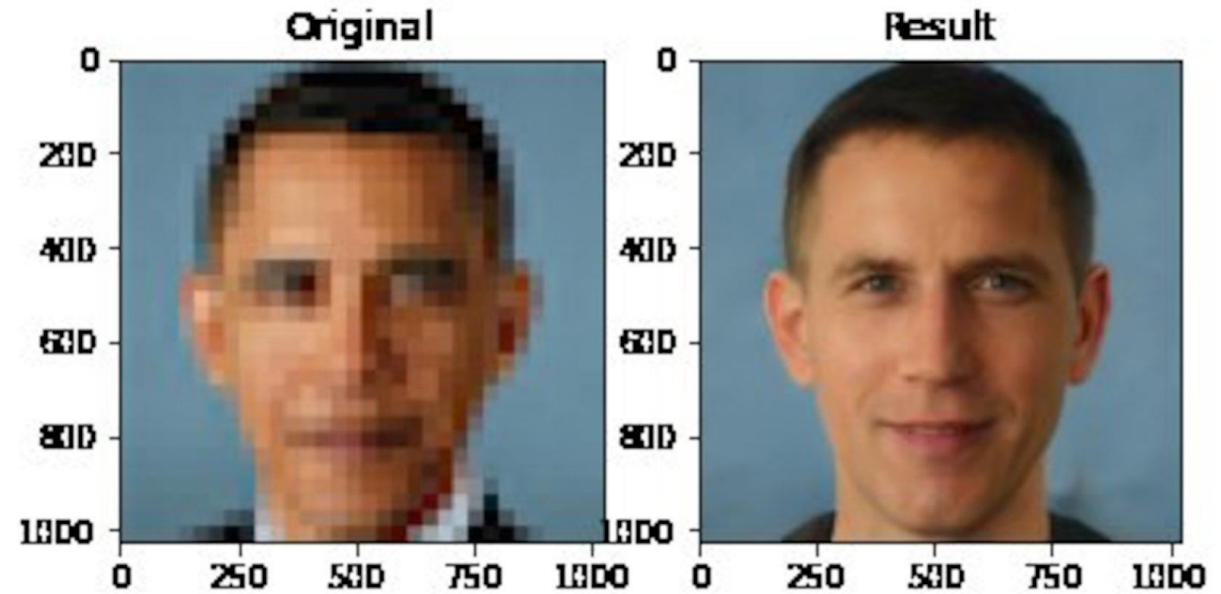
Uso cotidiano y potencial

Ejemplos sobre las aplicaciones actuales de la IA y posibilidades que ofrece

- Asistentes personales digitales en smartphone o PC
- Búsqueda en la web
- Traducción automática
- Ciberseguridad
- Lucha contra la desinformación
- Optimización de productos y procesos de ventas
- Aire acondicionado inteligente
- Vehículos autónomos
- Compras en línea y publicidad
- Internet de las cosas: aspiradoras, frigoríficos, relojes, etc., conectados a internet
- Agricultura inteligente: riego, alimentación de animales, robots anti malas hierbas
- Robots utilizados en fábricas

europarl.eu

Inteligencia Artificial – Ejemplos de Casos Fallidos



[Nota: 5 fallos recientes de la Inteligencia Artificial](#)

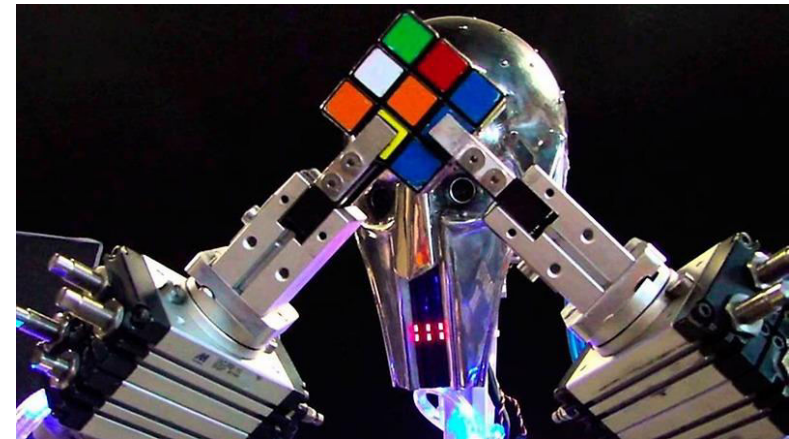
Inteligencia Artificial Débil

La Inteligencia Artificial débil trata de aquella de la cual ya conocemos y hacemos uso en nuestra cotidianidad.

Piensa en Siri de iOS, Amazon Alexa, los sistemas que permiten el hallazgo de información a través de distintas plataformas. Todos estos son ejemplos de inteligencia artificial débil.

¿Qué quiere decir esto?

Que se trata de aquella tecnología de la IA que hace uso de técnicas más o menos complejas que se centran en problemas concretos y que, más de allá, son incapaces de reemplazar a los humanos, pues no cuentan con el poder de razonamiento y conciencia propia.



Inteligencia Artificial Fuerte

La Inteligencia Artificial fuerte, por el contrario, es la que busca igualar o exceder la inteligencia humana promedio.

Este tipo de IA puede realizar con éxito cualquier tarea intelectual de cualquier ser humano.

Además de **poder realizar las funciones de los humanos**, este tipo de IA busca la construcción de máquinas (robots) que tengan la capacidad de pensar y ser conscientes de sí mismas y lo que ocurre a su alrededor.

Son capaces de tomar decisiones y sentir emociones.



El Hombre Bicentenario, con Robin Williams (1999)



Her (2013)

Inteligencia Artificial



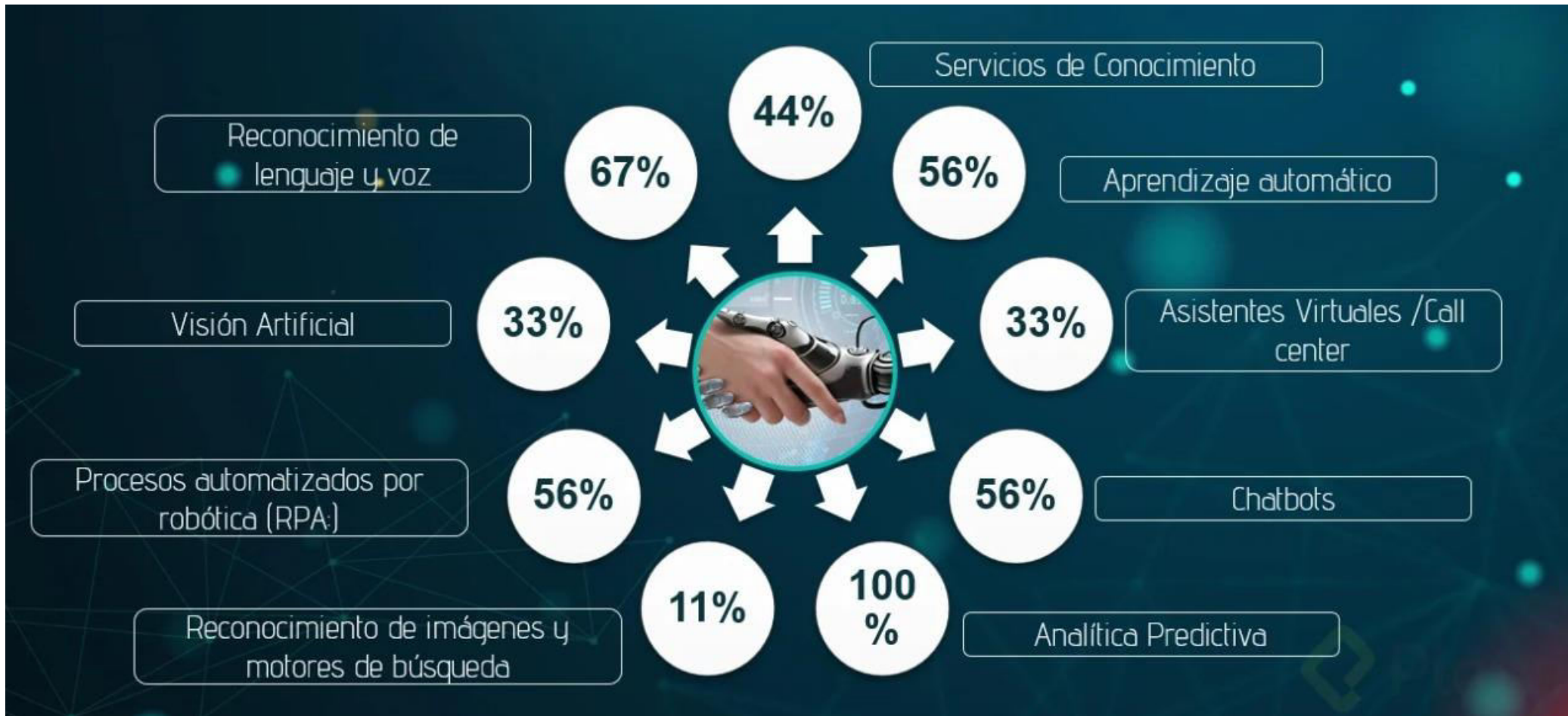
Fuerte:
Habilidades cognitivas similares a las humanas.

Débil:
Se desarrolla para realizar una tarea particular

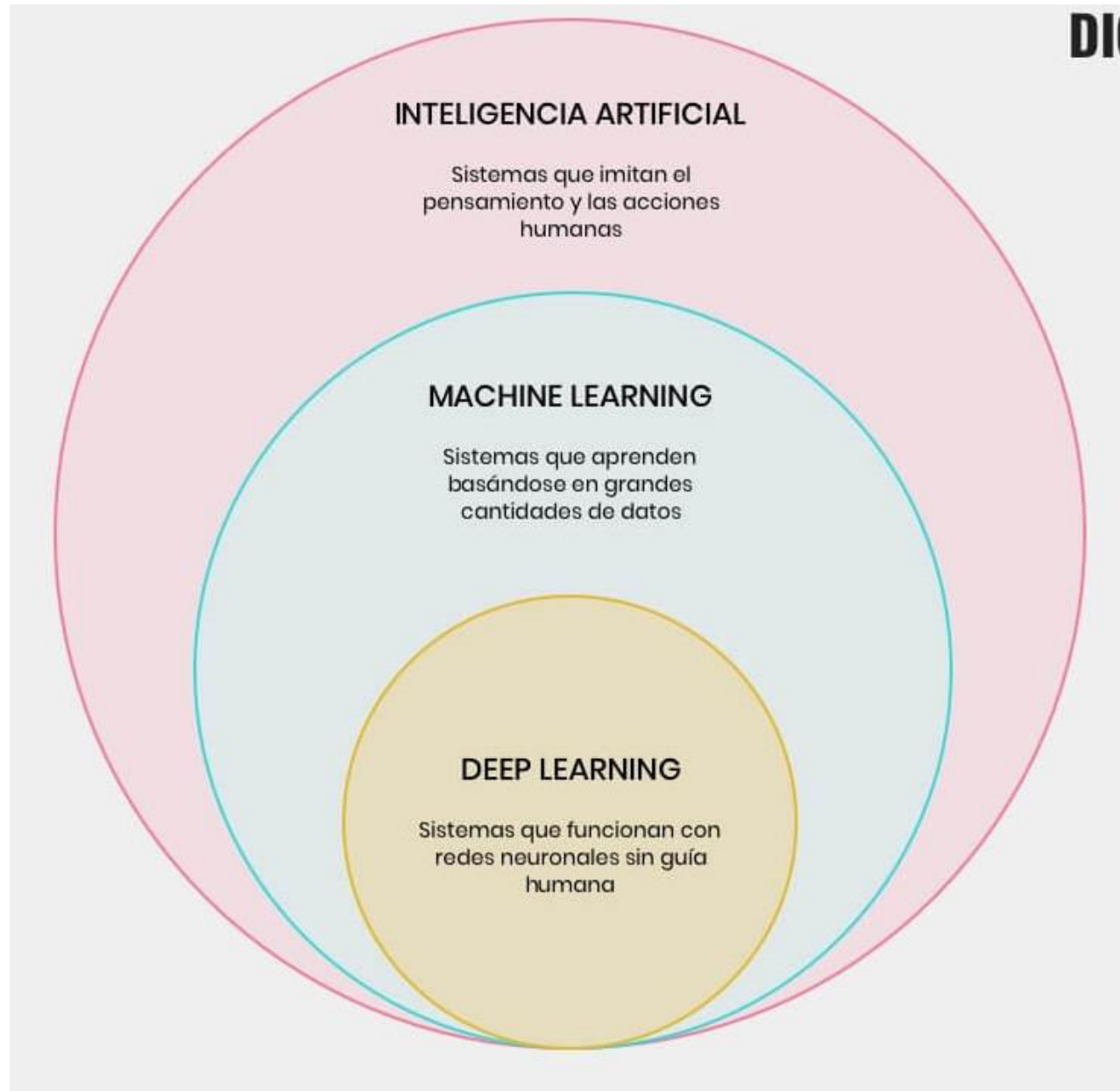
Ante tareas que no resultan ser familiares efectúa los cambios necesarios para cumplirlas.
(Machine Learning y Deep Learning)

Inteligencia Artificial

Algunas tecnologías de IA que se están aplicando hoy en día.



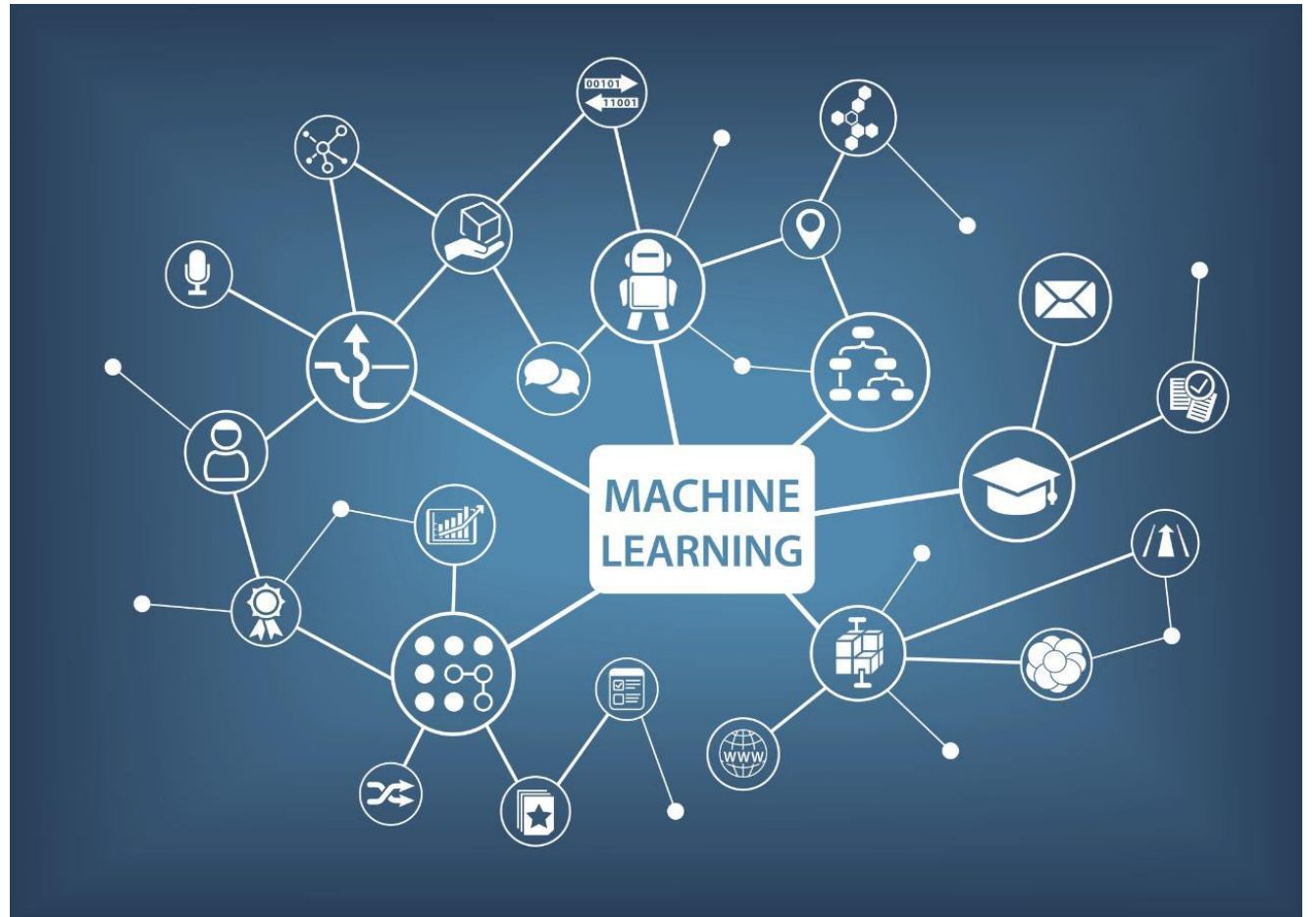
Inteligencia Artificial, Machine Learning y Deep Learning



Machine Learning

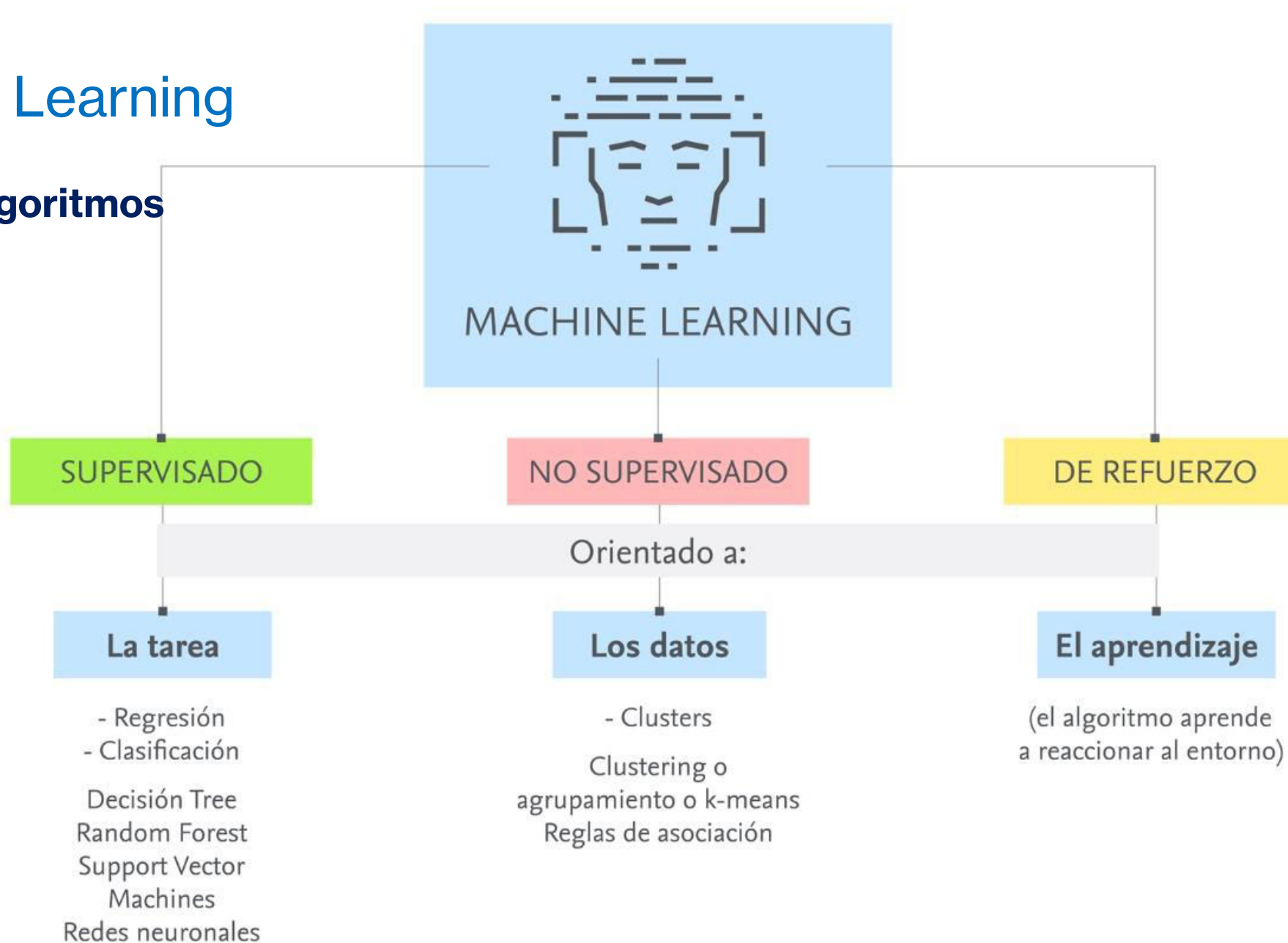
El **'machine learning'** –aprendizaje automático– es una rama de la inteligencia artificial que permite que las máquinas aprendan sin ser expresamente programadas para ello.

Una **habilidad indispensable** para hacer sistemas capaces de identificar patrones entre los datos para hacer predicciones.



Machine Learning

Tipos de Algoritmos



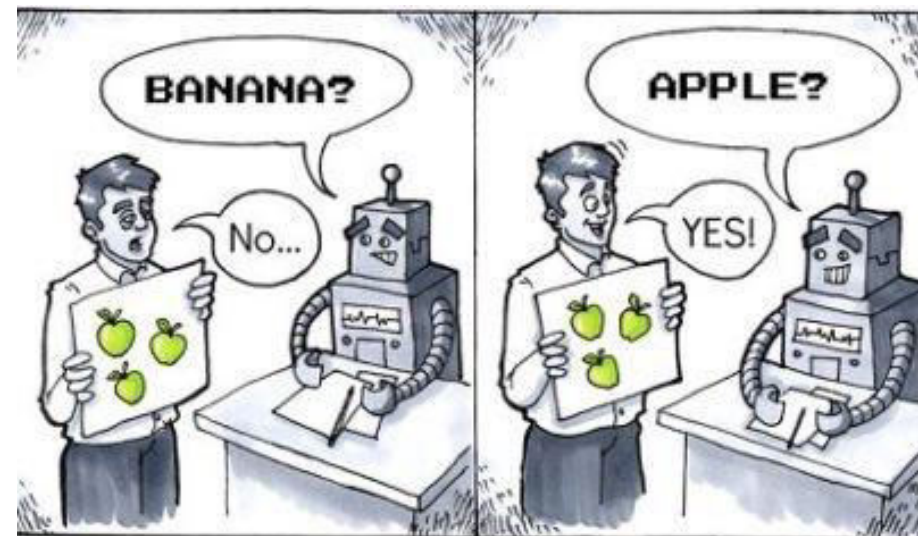
Machine Learning – Tipos de Aprendizaje

Aprendizaje supervisado

En el aprendizaje supervisado, la máquina se enseña con el ejemplo.

De este modo, el **operador** proporciona al algoritmo de aprendizaje automático un conjunto de datos conocidos que incluye las entradas y salidas deseadas, y **el algoritmo debe encontrar un método para determinar cómo llegar a esas entradas y salidas.**

Mientras el **operador conoce** las respuestas correctas al problema, el **algoritmo identifica patrones** en los datos, **aprende de las observaciones y hace predicciones.** El **algoritmo** realiza predicciones y es **corregido por el operador**, y este proceso sigue hasta que el algoritmo alcanza un alto nivel de **precisión y rendimiento.**



Supervised Learning

Machine Learning – Tipos de Aprendizaje

Aprendizaje **No** Supervisado

El algoritmo de **aprendizaje automático** estudia los datos para identificar patrones.

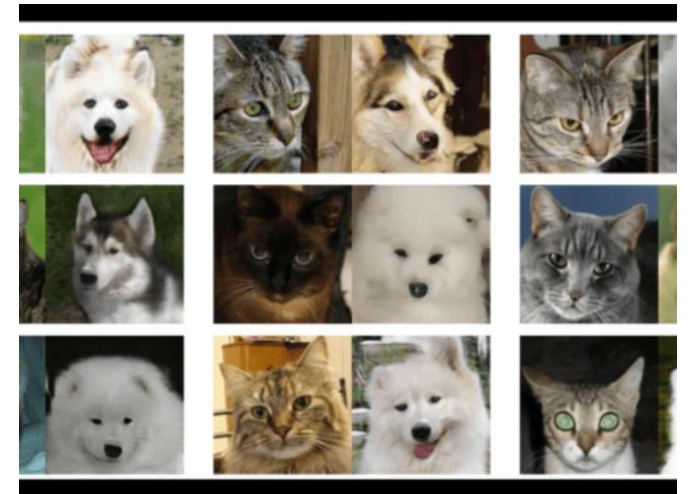
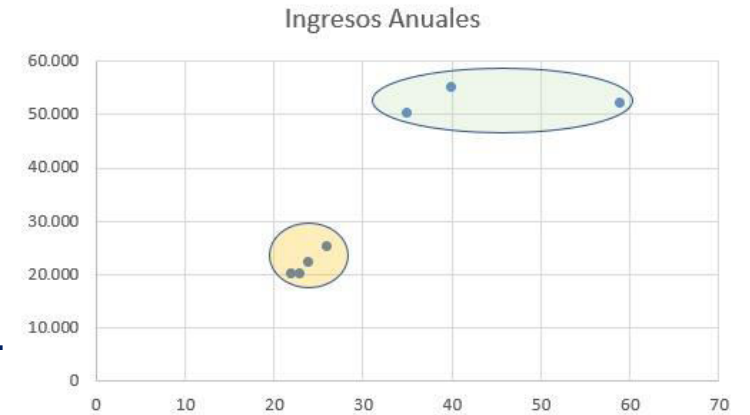
No hay una clave de respuesta o un operador humano para proporcionar instrucción.

En cambio, la **máquina determina** las correlaciones y las relaciones mediante el análisis de los datos disponibles.

En un proceso de aprendizaje no supervisado, **se deja que el algoritmo de aprendizaje automático** interprete grandes conjuntos de datos y dirija esos datos en consecuencia.

Así, el **algoritmo intenta organizar** esos datos de alguna manera para describir su estructura. Esto podría significar la necesidad de agrupar los datos en grupos u organizarlos de manera que se vean más organizados.

A medida que evalúa más datos, su capacidad para tomar decisiones sobre los mismos mejora gradualmente y se vuelve más refinada.



Machine Learning – Tipos de Aprendizaje

Aprendizaje por refuerzo

El aprendizaje por refuerzo **se centra en los procesos de aprendizajes reglamentados**, en los que se proporcionan algoritmos de aprendizaje automáticos con un conjunto de acciones, parámetros y valores finales.

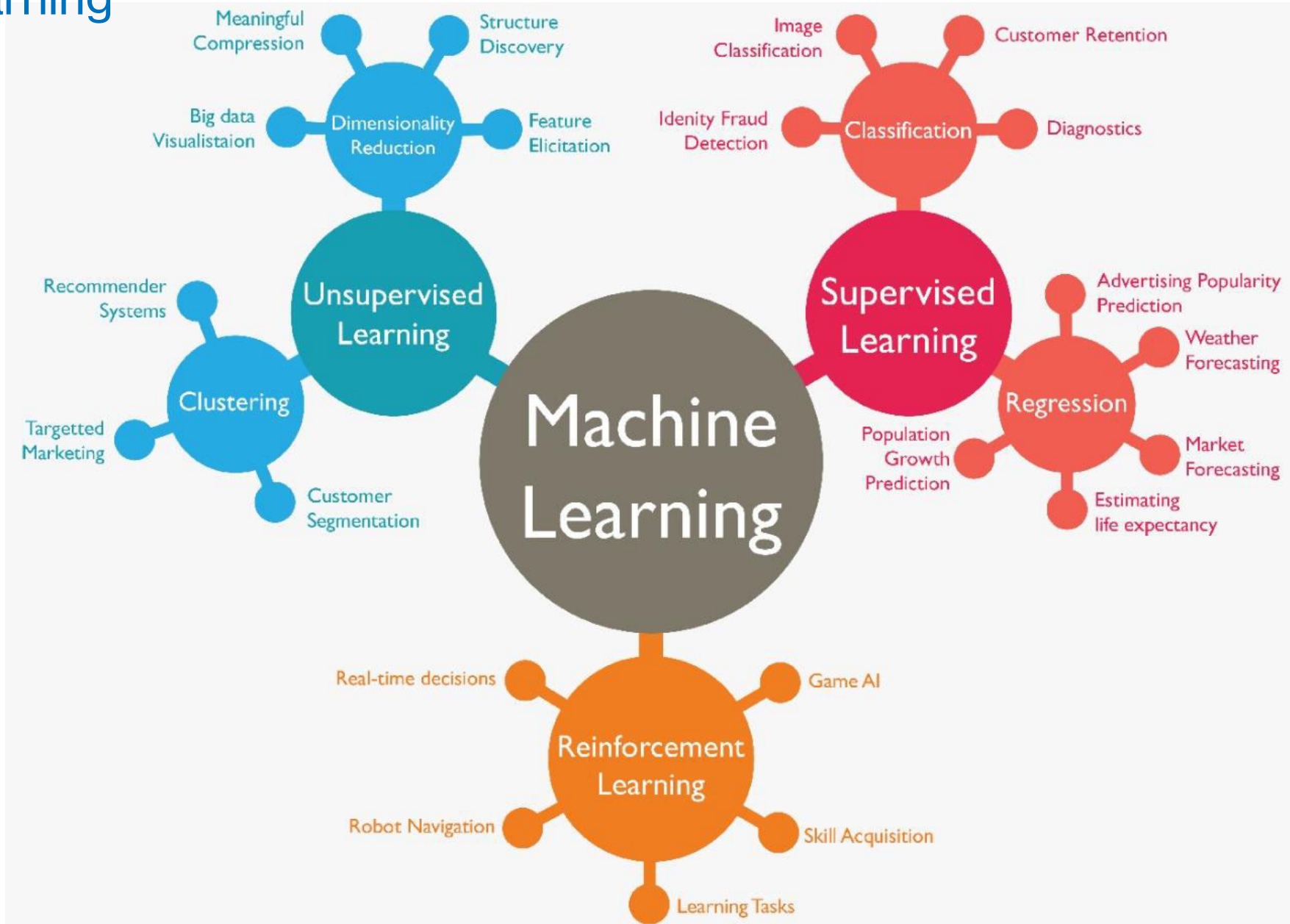
Al definir las reglas, el **algoritmo** de aprendizaje automático **intenta explorar diferentes opciones y posibilidades**, monitorizando y evaluando cada resultado para determinar cuál es el óptimo.

En **consecuencia**, este sistema enseña la máquina a través del proceso de ensayo y error.

Aprende de experiencias pasadas y comienza a adaptar su enfoque en respuesta a la situación para lograr el mejor resultado posible.



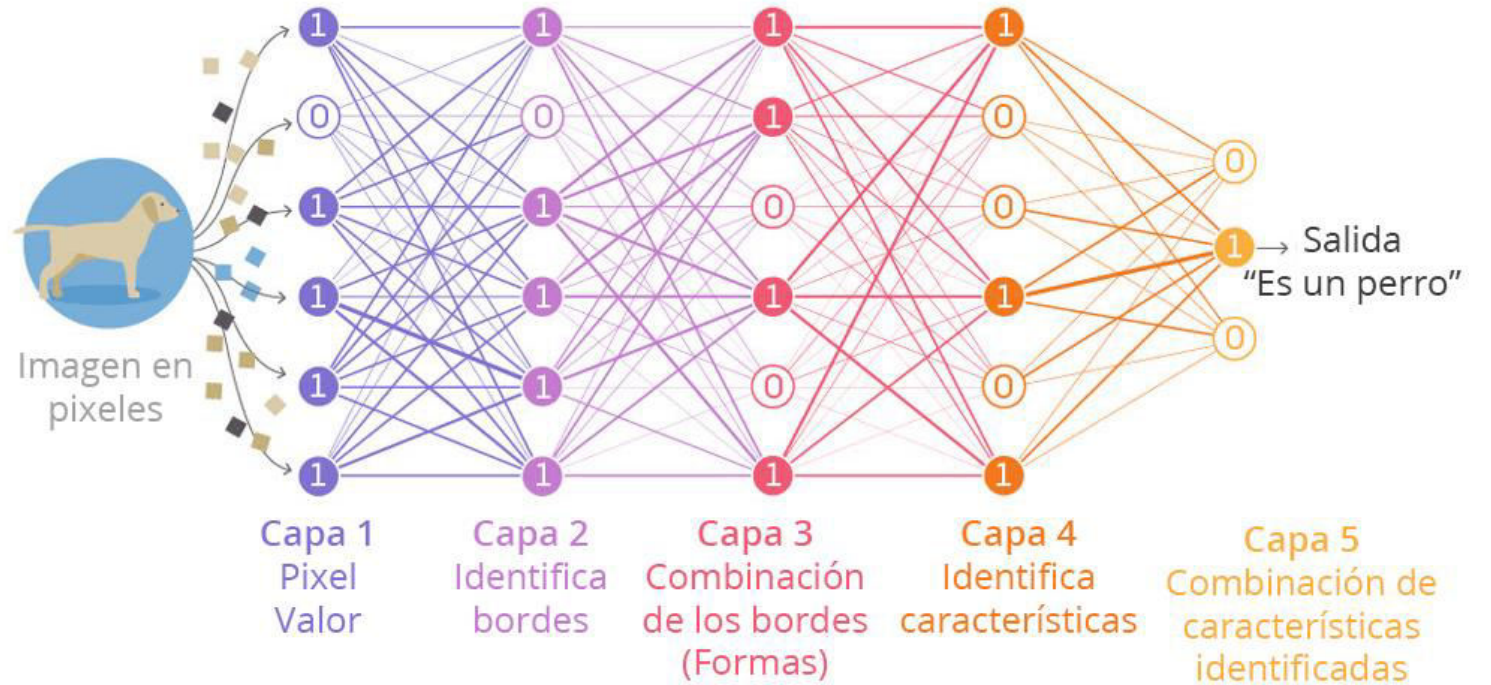
Machine Learning



Deep Learning

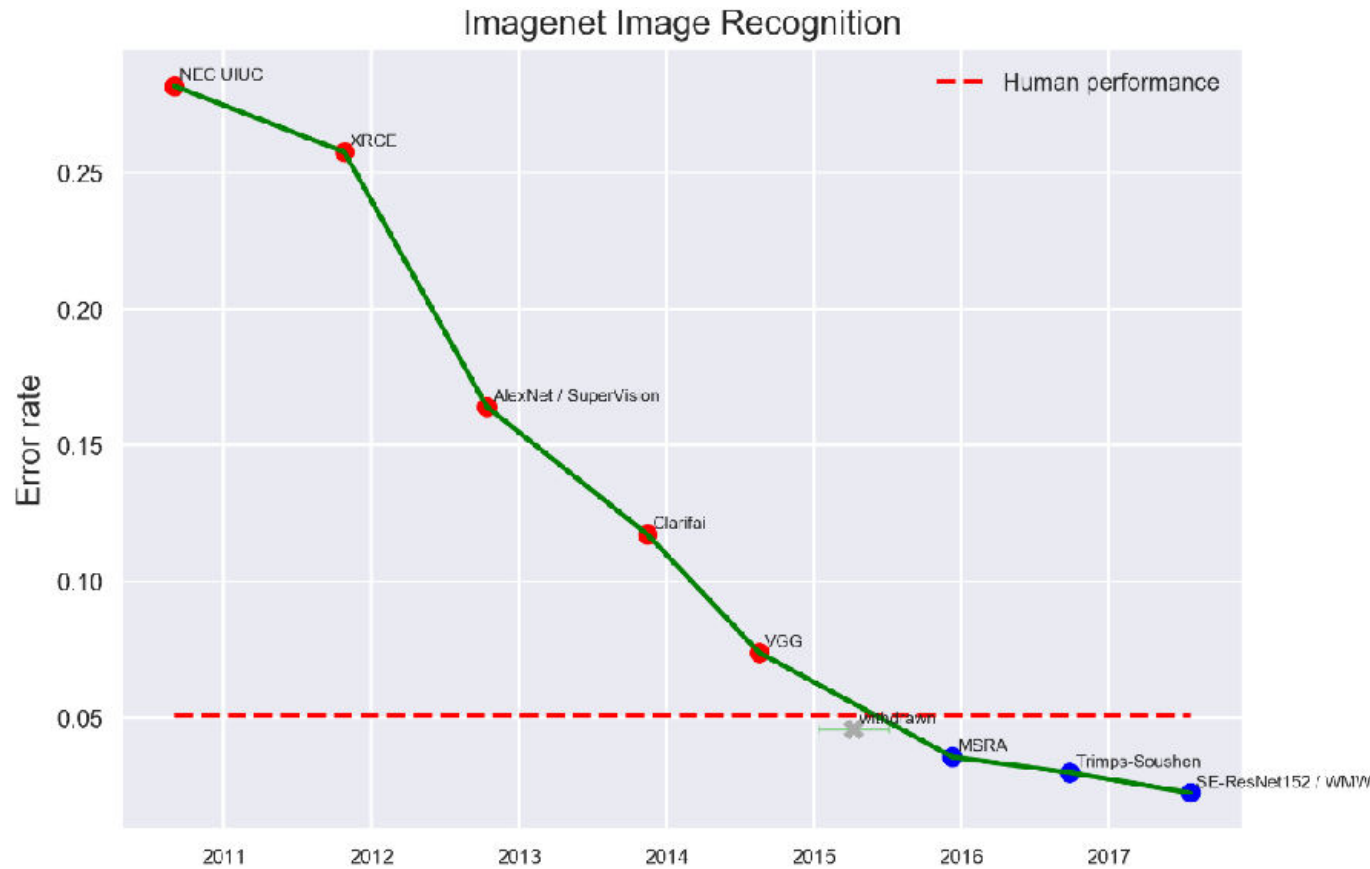
Deep Learning

Es un subcampo específico del machine learning: una nueva forma de aprender representaciones a partir de datos que pone énfasis en el aprendizaje de capas sucesivas de representaciones cada vez más significativas.



Fuente: <https://www.quantamagazine.org/>

Deep Learning



Los modelos computacionales de Deep Learning **imitan** las características arquitecturales del sistema nervioso humano, permitiendo que dentro del sistema global haya redes de unidades de proceso que se especialicen en la detección de determinadas características ocultas en los datos.

Deep Learning

En marzo de 2016, AlphaGo de Deepmind venció al jugador de Go 18 veces campeón mundial Lee Sedol 4-1 en una serie vista por más de 200 millones de personas.

Una máquina había aprendido una estrategia sobrehumana para jugar al Go, una hazaña que antes se creía imposible, o al menos, al menos a una década de realizarse.



[Nota: Una victoria de la IA que desnuda los secretos de la mente](#)

Deep Learning



Esto en sí mismo fue un logro notable. Sin embargo, el 18 de octubre de 2017, DeepMind dio un gran paso más allá.

El artículo " **Dominando el juego del Go sin conocimiento humano** " reveló una nueva variante del algoritmo, AlphaGo Zero, que había derrotado a AlphaGo 100-0.

Increíblemente, lo había hecho aprendiendo únicamente a través del autojuego y encontrando gradualmente estrategias que superarían encarnaciones anteriores de sí mismo.

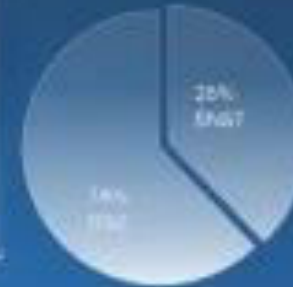
Ya no se necesitaba una base de datos de juegos de expertos humanos para construir una IA sobrehumana.

Data Analytics

Projected sales of main products in 2013



Distribution of market share among the major industry players



Distribution of market share among the major industry players: IT & C and B & A take 14% and 56% percent respectively. A further change in the economic situation of the market will be characterized by a more equal distribution of market share among players.

Share of market activity



Changes in the activity of the active and passive market is uncertain. Established positive trends in various market segments.

Projected sales of main products in 2013



Passive market share

Data Analytics

[Video: ¿Qué es Big Data Analytics?](#)

¿Por qué es importante conocer los tipos de Data Analytics?

Con varios tipos de **business analytics**, las empresas son libres de elegir la profundidad a la que deben sumergirse en el análisis de datos.

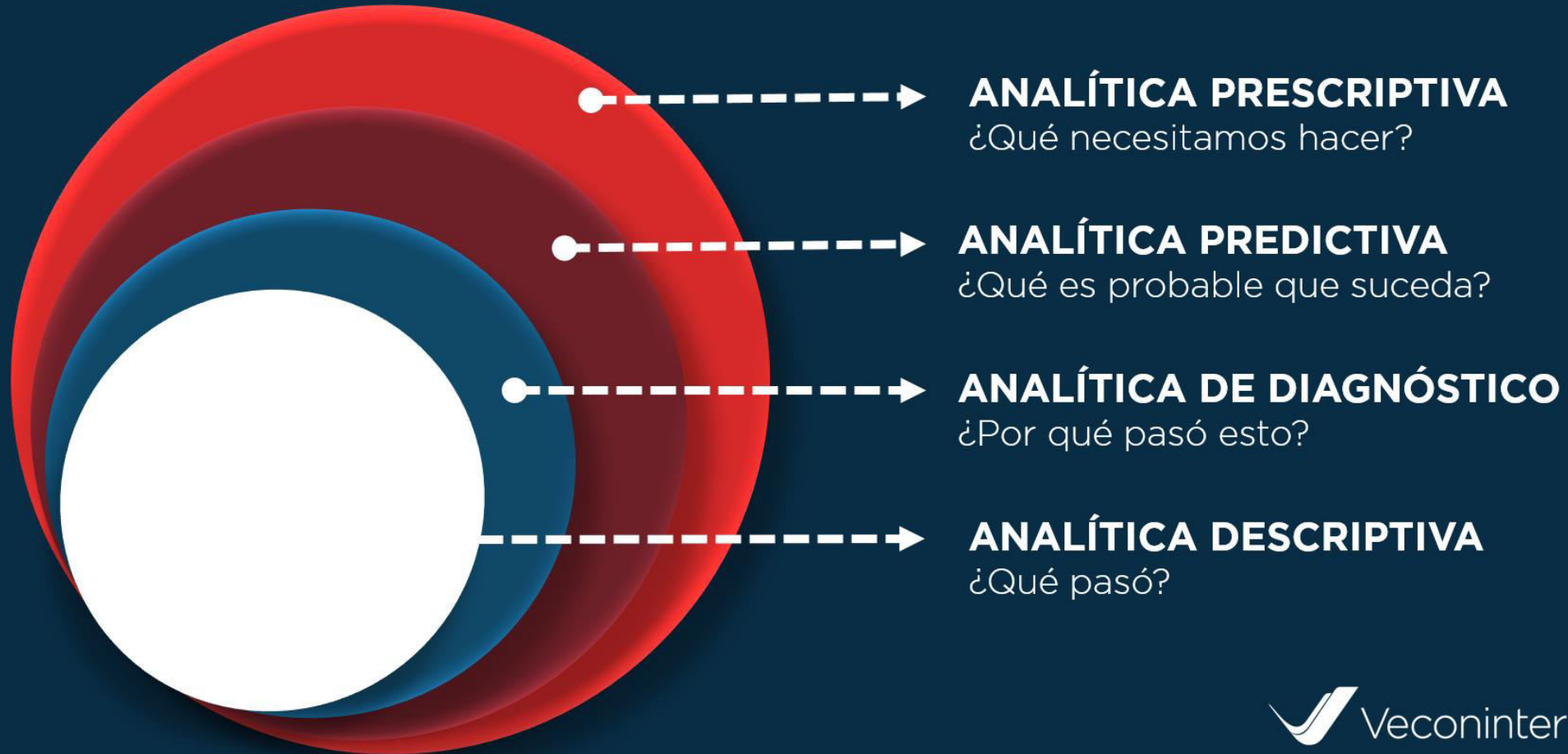
El análisis **descriptivo** es fundamental para la toma de decisiones ágiles en las unidades de negocio, y las empresas deben buscar ser sólidas en esta área al mismo tiempo que desarrollan habilidades en campos más sofisticados de business analytics.

Por otro lado, el análisis **predictivo** y **prescriptivo** hace que las empresas sean proactivas en el uso de los datos y en muchos casos permiten generar ventajas competitivas.



Data Analytics

LOS 4 TIPOS DE **DATA ANALYTICS**



Data Analytics

Analítica Descriptiva

El análisis descriptivo es una etapa preliminar del tratamiento de datos que consiste en sintetizar los datos históricos para obtener información útil o incluso prepararlos para un análisis posterior.

Este análisis se utiliza para buscar y **resumir datos históricos** con el fin de identificar patrones o significados.

El análisis descriptivo consiste en tratar de describir o resumir los datos. Aunque **no hace predicciones sobre el futuro**, puede ser muy valioso en los entornos empresariales. Esto se debe principalmente a que el análisis tipo descriptivo facilita el consumo de datos, lo que puede facilitar la actuación de los analistas.

Se centra únicamente en los datos históricos.



Por medio de los datos ayudamos a las empresas a comprender cómo se están desempeñando y responder la pregunta ¿Qué pasó?. Esto por medio de Inteligencia de Negocios, visualización de datos, gráficos, cuadros, informes y paneles.

Data Analytics

Analítica de Diagnostico

La analítica diagnóstica también es **retrospectiva**, pero en cambio, **busca el por qué ocurrió el problema** que se presentó en el análisis descriptivo y es el siguiente paso esencial.

El análisis de diagnóstico también aprovecha los datos históricos de una empresa proveniente de diversas fuentes internas.

Este análisis **es más complejo** y requiere que los analistas de datos utilicen estructuras de datos multidimensionales, algoritmos estadísticos y capacidades de slice-and-dice, drill-through, drill-down, roll-up, etc. para poder encontrar fácil y rápidamente las causas de algún problema, patrones, tendencias y correlaciones.

La **mayor ventaja** de las analíticas de diagnóstico es poder **proporcionar contexto a un problema empresarial** a través de una serie de modelos de datos.



Con la analítica de diagnóstico vamos un paso más allá y proporcionamos un análisis profundo para responder la pregunta ¿Por qué pasa esto?. Esto lo logramos mediante la aplicación de minería de datos.

Data Analytics

Analítica de Predictiva



Utilizando la información histórica y actual hacemos predicciones acerca del futuro para ayudar a las empresas a identificar riesgos y oportunidades. Contestar la pregunta ¿Qué podría pasar si?. Para este fin aplicamos métodos de Machine Learning.

El análisis predictivo es una forma de **análisis avanzado** que utiliza datos nuevos e históricos para **pronosticar** la actividad, el comportamiento y las **tendencias**.

Implica aplicar **técnicas de análisis estadístico**, consultas analíticas y algoritmos de aprendizaje automático para crear modelos predictivos que otorguen un valor numérico – o puntuación – a la probabilidad de que ocurra un evento en particular. Por ejemplo para **predecir el comportamiento** probable de las personas, la maquinaria u otras entidades.

Se puede utilizar para una variedad de casos de uso. Por ejemplo, es probable que una compañía de seguros tenga en cuenta las posibles variables de seguridad en la conducción, como la edad, el sexo, la ubicación, el tipo de vehículo y el historial de conducción, al fijar los precios y emitir pólizas de seguro de automóvil.

Data Analytics

Analítica Prescriptiva



Abordamos la pregunta ¿qué se debe hacer?. Sugerimos varios cursos de acción y describimos las posibles implicaciones para cada uno. Esto lo logramos utilizando técnicas de optimización de decisiones.

La analítica **predice lo que ocurrirá, cuándo ocurrirá y las razones por las que ocurrirá**. Su trabajo consiste en sintetizar de forma automática grandes datos, ciencias matemáticas, reglas de negocio y aprendizaje automático para hacer predicciones y sugerir opciones que ayuden a decidir.

Va más allá de predecir los resultados futuros, ya que también **sugiere acciones** para beneficiarse de las predicciones y mostrar al tomador de decisiones las implicaciones de cada opción.

El **análisis prescriptivo** se encarga de encontrar una **solución entre una gama de variantes** con el objetivo de optimizar los recursos y aumentar la eficiencia operativa.

Esta herramienta **utiliza** diferentes **técnicas de simulación y optimización** para señalar el camino que realmente conviene seguir.

Data Analytics - Comparativo

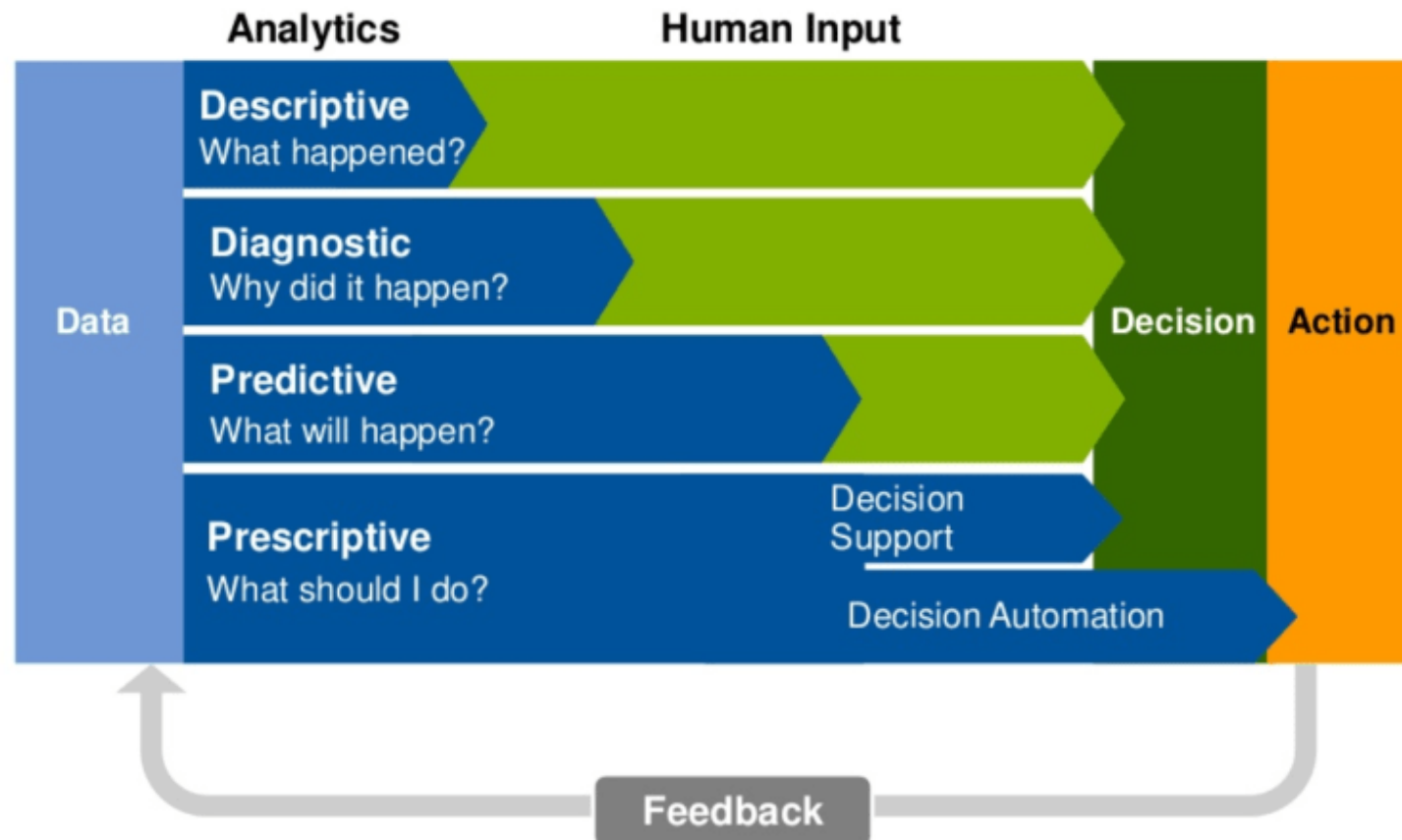
Evolución del análisis



Data Analytics

Hasta donde llega el ser Humano

The Analytics Continuum





BUSINESS INTELLIGENCE

Business Intelligence

Datos Vs. Información Vs. Conocimiento

DATO

Es un **valor**...
Por ejemplo, “500 dólares”.

INFORMACIÓN

Tiene un **contexto**...
“Las ventas del mayo fueron de 500 dólares”

CONOCIMIENTO

Se obtiene mediante el **análisis** de la información...
“Mayo es el mes más alto en ventas de este año”

[Video: El ABC del Business Intelligence: estos conceptos no significan lo mismo](#)

Business Intelligence

“Conjunto de procesos y herramientas orientadas al análisis de información con el objetivo de hacer uso de datos reales en el proceso de toma de decisiones.”



El término **Business Intelligence** (BI por sus siglas en inglés) hace referencia al uso de estrategias y herramientas que sirven para transformar información en conocimiento, con el objetivo de mejorar el proceso de toma de decisiones en una empresa.

En plena era digital, tomar **decisiones bien informadas** es uno de los principales factores de diferenciación de las empresas.

Business Intelligence

Hoy en día las empresas tienen a su alcance cantidades inimaginables de datos, los cuales pueden generarse internamente u obtenerse desde fuentes externas, la pregunta es:

¿qué hacen con los datos?

¿Obtienen algún beneficio de ellos?

¿Se consideran para la toma de decisiones?

Justamente para responder estos cuestionamientos está la **Inteligencia de Negocios** o **Business Intelligence (BI)**

Para esto se usa software especializados en Inteligencia de Negocios los cuales contribuyen a que esta **visión** y **estrategia** pueda ocurrir, es a través de ellos que se pueden concentrar, visualizar y analizar los datos.

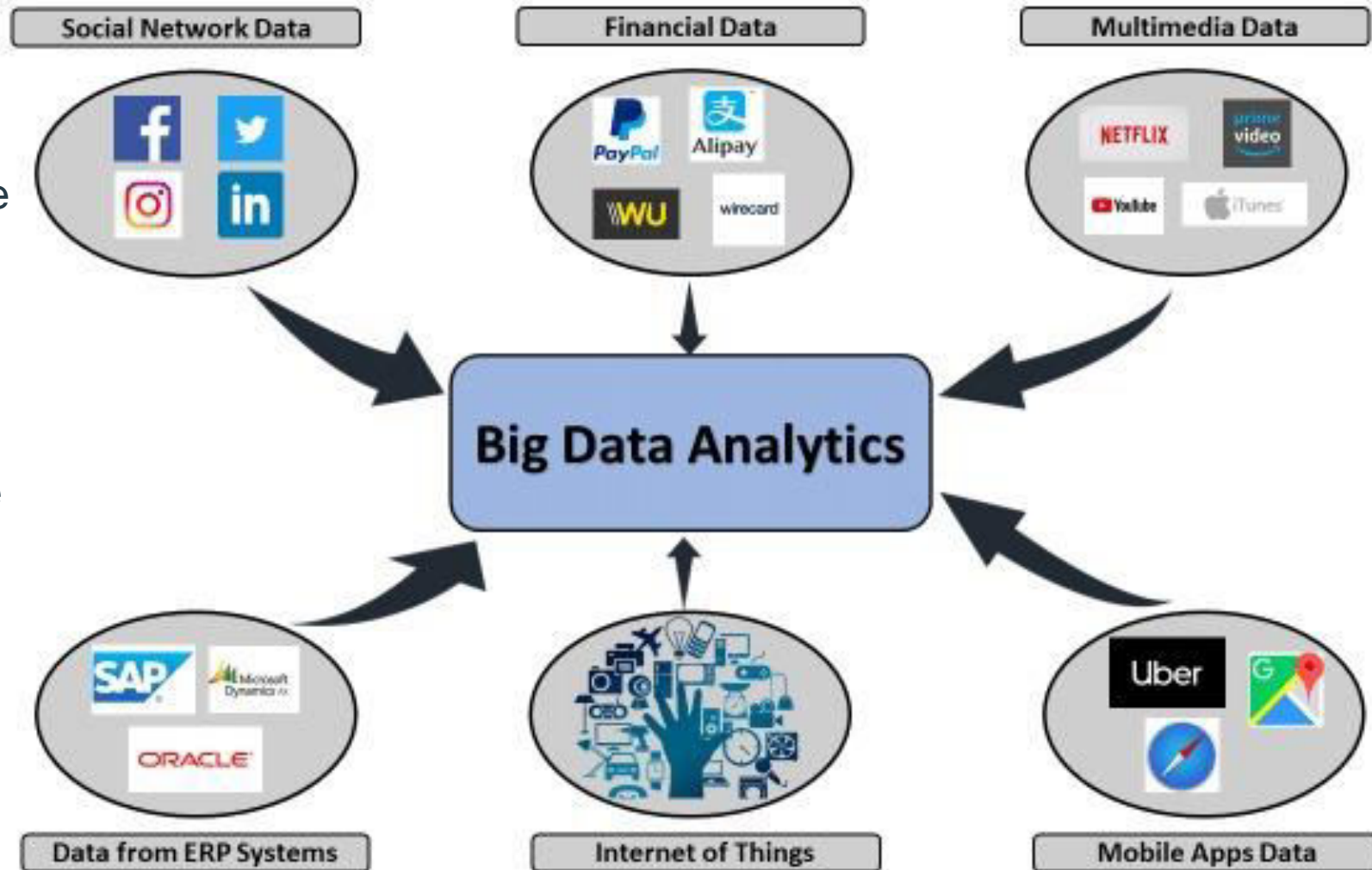
Algunos ejemplos de herramientas para análisis de datos son: Power BI, Tableau, Qlik, Microstrategy.



Business Intelligence

Fuentes de Datos

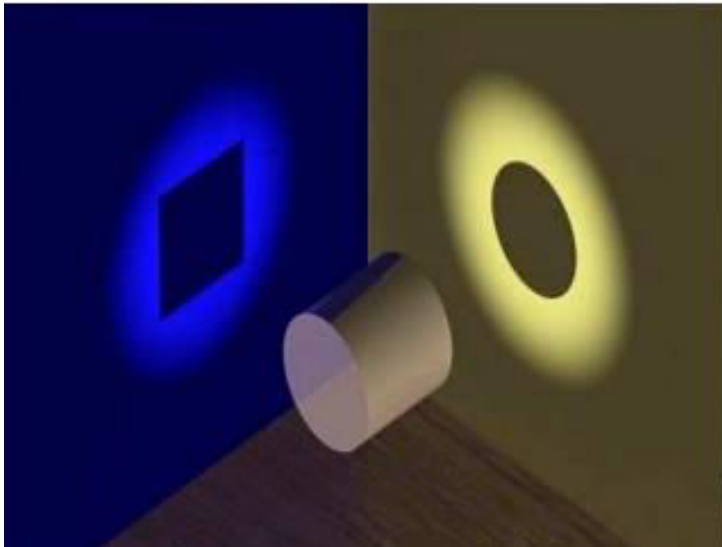
En cuanto a estos datos que tanto se mencionan en Business Intelligence, surge la pregunta respecto a las fuentes de las que se obtienen son:



Business Intelligence

Desafíos que hay que resolver:

Visión Integral



Acceso a la información necesaria en el tiempo adecuado



Única Verdad



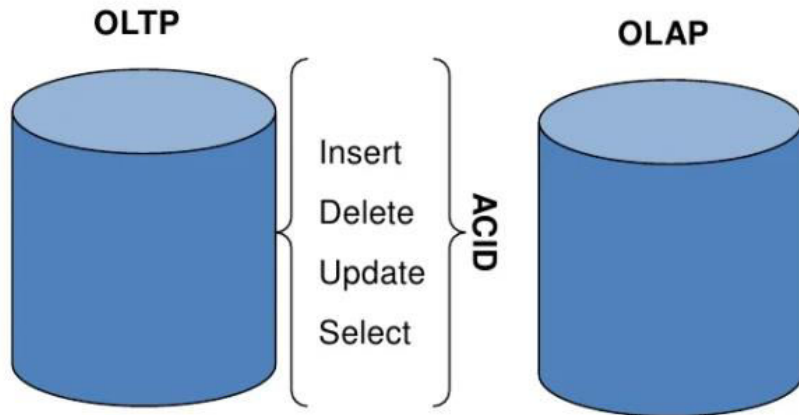
Business Intelligence



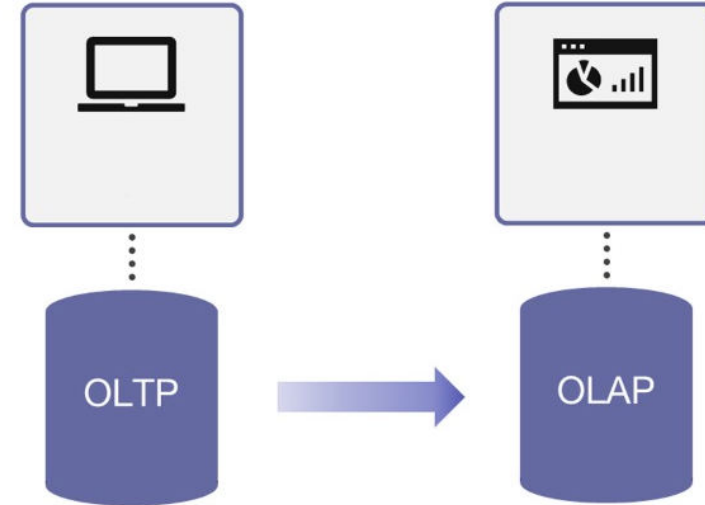
Business Intelligence

Tipos de Sistemas

- OLTP (On Line **T**ransaction Processing)
(Procesamiento de transacción en línea)
Bases de datos orientadas al procesamiento de transacciones.
- OLAP (On Line **A**nalytic Processing)
(Procesamiento analítico en línea)
Bases de datos orientadas al procesamiento analítico.



Alto Volumen de transacciones.
Procesamiento rápido.
Muchas tablas



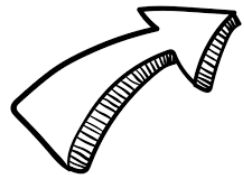
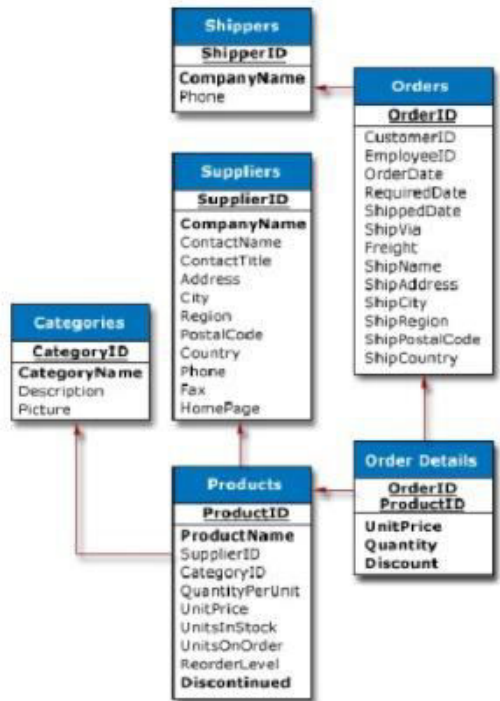
Alto volumen de datos
consultas lentas

Ejemplo de **OLTP**: operaciones de supermercado, reservas aéreas, operaciones bancarias.

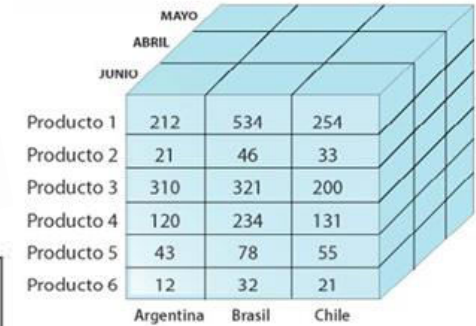
El **ejemplo** típico de cubo **OLAP** consiste invariablemente en un cubo comercial con las ventas en unidades e importes. Así, tenemos las ventas por producto, por familia, por categoría, o por fechas, semanas, meses, años, o por cliente, provincia, región o país.

Business Intelligence

Tipos de Sistemas

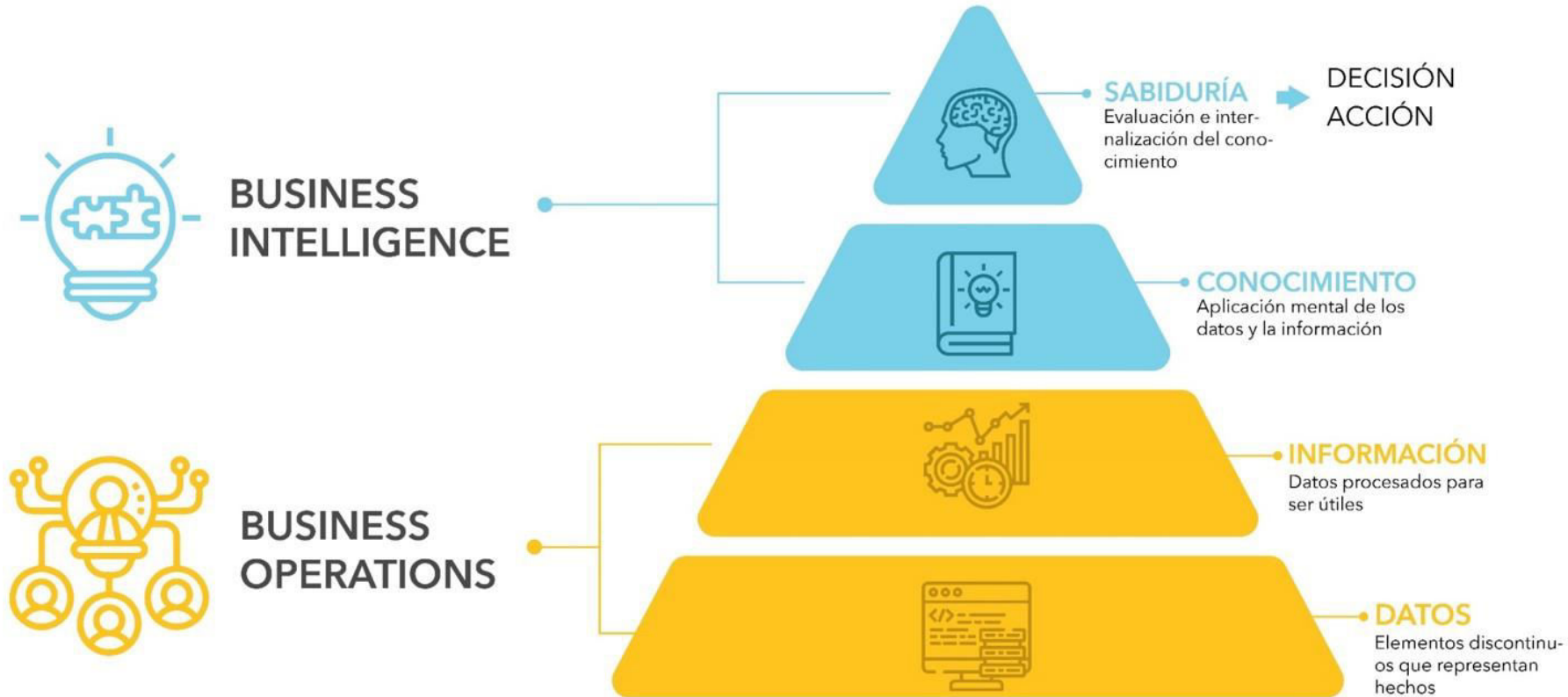


Sistema Operacional (OLTP)	Data Warehouse (Basado en Modelos Dimensionales: OLAP)
Almacena datos actuales.	Almacena datos históricos.
Almacena datos de detalle.	Almacena datos de detalle y datos agregados a distintos niveles.
Los datos son dinámicos (actualizables).	Los datos son estáticos.
Las transacciones son repetitivas.	Los procesos no son previsibles.
El número de transacciones es elevado.	El número de transacciones es bajo o medio.
Dedicado al procesamiento de transacciones.	Dedicado al análisis de datos.
Orientado a los procesos de la organización.	Orientado a la información relevante.
Soporta decisiones diarias.	Soporta decisiones estratégicas.
Sirve a muchos usuarios administrativos.	Sirve a técnicos de dirección.



Business Intelligence

La pirámide de la información



Fuente: Hey, J.: The Data, Information, Knowledge, Wisdom Chaim: The Metaphorical Link

Business Intelligence

BI A NIVEL ORGANIZACIONAL



TP N° 5: Sistemas de Información



Marque la diferencia entre los sistemas OLTP y OLAP.

-Traer un ejemplo de al menos 1 de ellos para compartir en la próxima clase.

Armar una presentación en 1 hoja en base a lo solicitado para exponer en clase

Éxitos!!

TP N° 6: Visualización de datos



Listar al menos 2 herramientas para la visualización de datos.

-Incluir las ventajas y desventajas de cada una.

Armar una presentación en 1 hoja en base a lo solicitado para exponer en clase.

Éxitos!!

BIG DATA & ANALYTICS I



TEMARIO

- Módulo 1: Surgimiento de Big Data
- Módulo 2: Los Datos Nos Rodean
- Módulo 3: Obtención de los Datos y Proceso de Toma de Decisión
- Módulo 4: Visualizar con Datos

Disertantes: Lic. Maria Trinidad Aquino – Ing. Raúl Alejandro Grassi

CePETel

Sindicato de los Profesionales
de las Telecomunicaciones

SECRETARÍA TÉCNICA



Instituto Profesional de
Estudios e Investigación

BIG DATA & ANALYTICS I

Módulo 4: Visualizar con Datos

Business Intelligence

“Conjunto de procesos y herramientas orientadas al análisis de información con el objetivo de hacer uso de datos reales en el proceso de toma de decisiones.”



El término **Business Intelligence** (BI por sus siglas en inglés) hace referencia al uso de estrategias y herramientas que sirven para transformar información en conocimiento, con el objetivo de mejorar el proceso de toma de decisiones en una empresa.

En plena era digital, tomar **decisiones bien informadas** es uno de los principales factores de diferenciación de las empresas.

Business Intelligence

Datos Vs. Información Vs. Conocimiento

DATO

Es un **valor...**
Por ejemplo, “500 dólares”.

INFORMACIÓN

Tiene un **contexto...**
“Las ventas del mayo fueron de 500 dólares”

CONOCIMIENTO

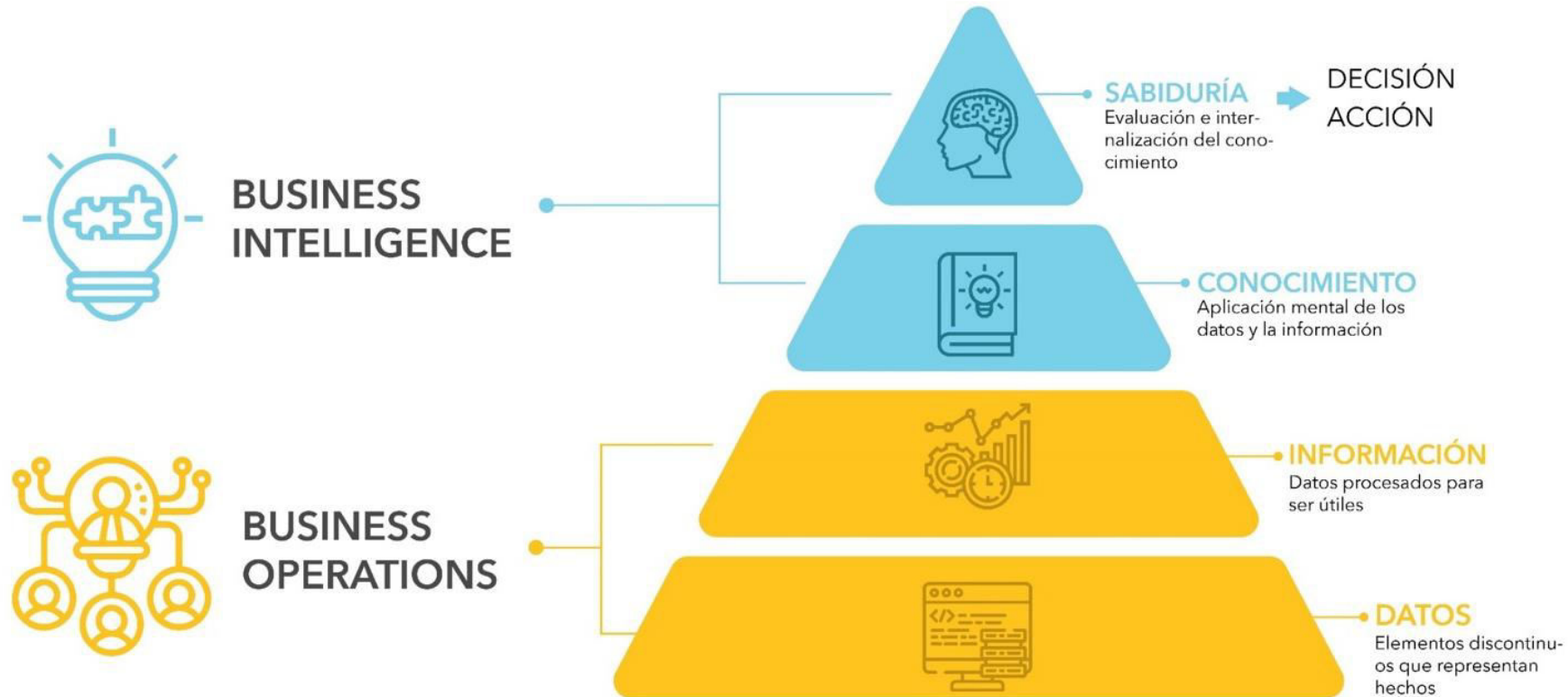
Se obtiene mediante el **análisis** de la información...
“Mayo es el mes más alto en ventas de este año”

[Video: El ABC del Business Intelligence: estos conceptos no significan lo mismo](#)

Business Intelligence

[Video: ¿Qué es visualización de datos?](#)

La pirámide de la información



Fuente: Hey, J.: The Data, Information, Knowledge, Wisdom Chain: The Metaphorical Link

Toma de decisiones

¿Como se toman las decisiones?

Intuición vs. Realidad?

Realidad implica...

1. Identificar el problema
2. Encontrar alternativas
3. Evaluar resultados de cada una



Información y conocimiento forman la base del proceso de toma de decisiones



Sistemas de Información

“Un sistema de información (SI) es un **conjunto de elementos** orientados al tratamiento y administración de datos e información, organizados y listos para su uso posterior, **generados para cubrir una necesidad o un objetivo.**” – Wikipedia 2019

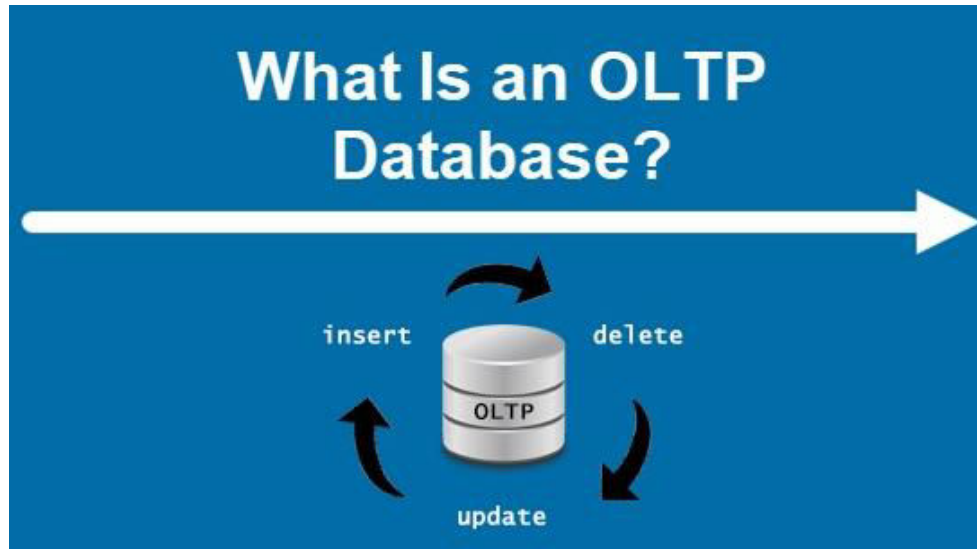


Sistemas de Información



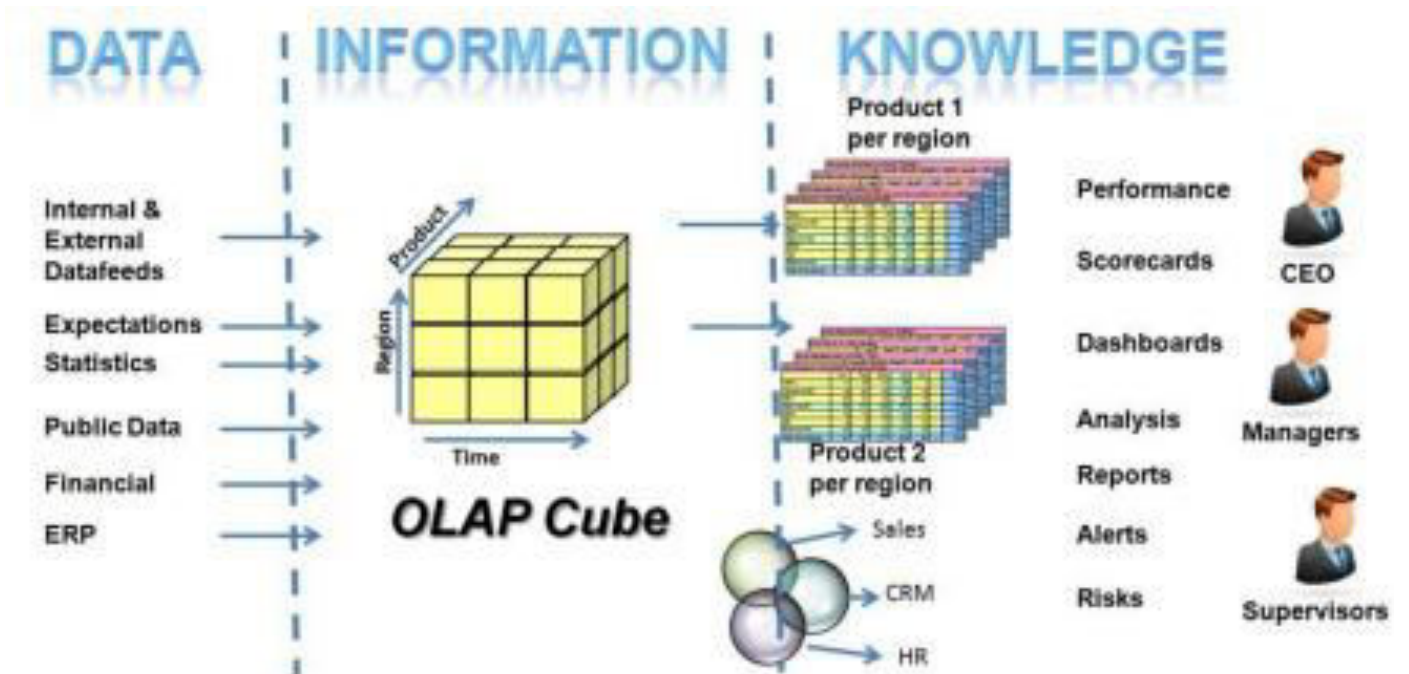
OLTP - On-Line Transactional Processing

- Bases de datos orientadas al procesamiento de transacciones.
 - Una transacción genera un proceso atómico, puede involucrar operaciones de inserción, modificación y borrado de datos.
 - El proceso transaccional es típico de las bases de datos operacionales
 - El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura.
 - Los datos se estructuran según el nivel aplicación (programa de gestión a medida, ERP o CRM implantado).
- Los formatos de los datos no son necesariamente uniformes en los diferentes departamentos



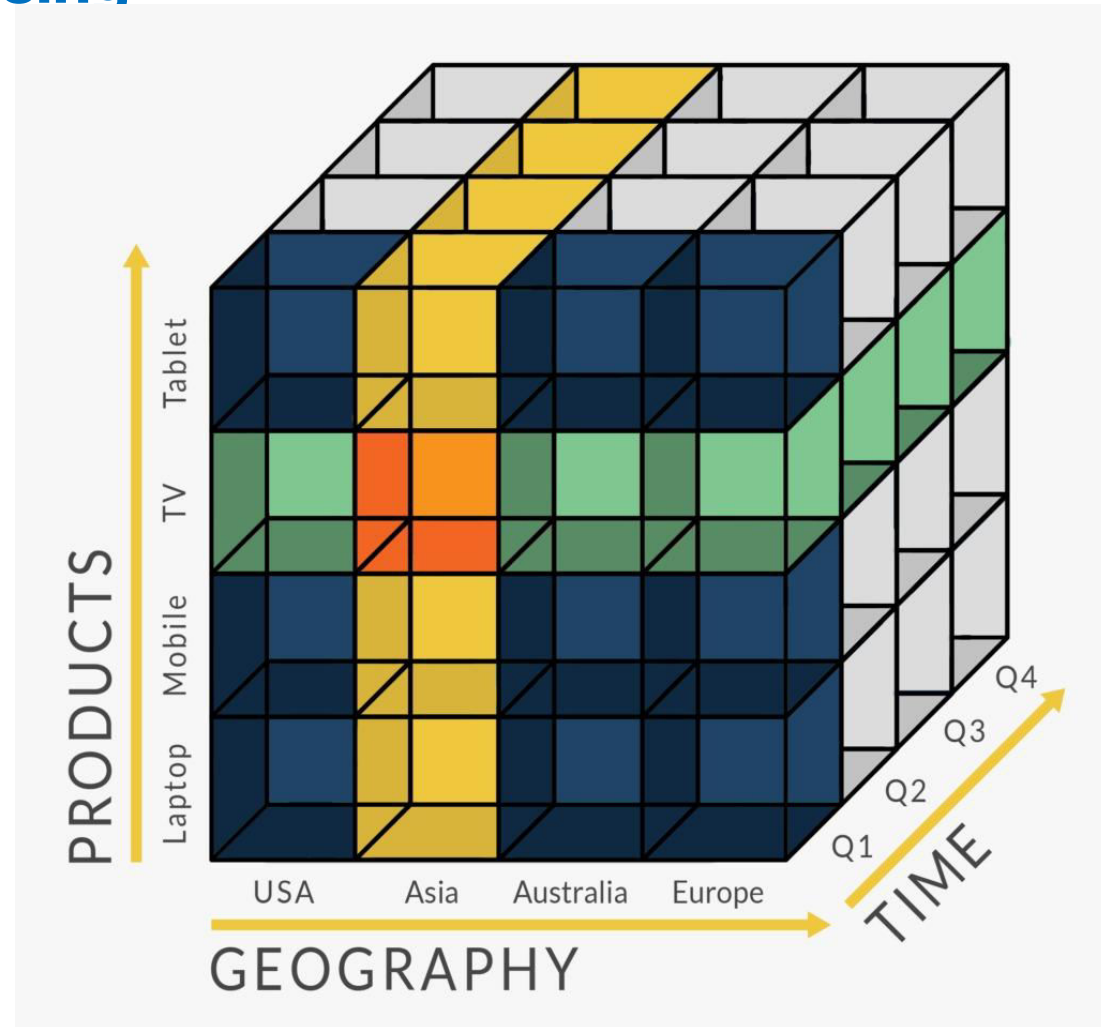
OLAP - On-Line Analytical Processing

- Bases de datos orientadas **al procesamiento analítico**. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a **extraer algún tipo de información útil**: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos, etc.
- Acceso a los datos de **sólo lectura**. La acción más común es la consulta.



OLAP - On-Line Analytical Processing

- Los datos se estructuran según las áreas de negocio, y los **formatos de los datos** están integrados de manera **uniforme** en toda la organización.
- El historial de datos es a **largo plazo**, normalmente de dos a cinco años.
- Las bases de datos OLAP se **suelen alimentar de información procedente de los sistemas operacionales** existentes, mediante un proceso de **extracción, transformación y carga (ETL)**.



Sistemas OLTP vs OLAP

OLTP	OLAP
Alineados por aplicación o funcionalidad. Poca integración, fronteras tecnológicas definidas.	Integrados. Abarcan varios procesos de negocio, u cruzan información entre si.
Actualización online .	Actualización batch .
Acceso transaccional (<i>create, read, update, delete</i>).	Acceso de solo lectura (<i>read</i>).
Datos recientes o de periodos de tiempos cortos.	Datos Históricos.
Información detallada.	Información agregada.
Normalización / no redundante.	Desnormalización / redundante.
Favorecer la Operación Transaccional del día a día .	Favorecer el Análisis .

Niveles de realización de BI

Análisis OLAP/Multidimensional



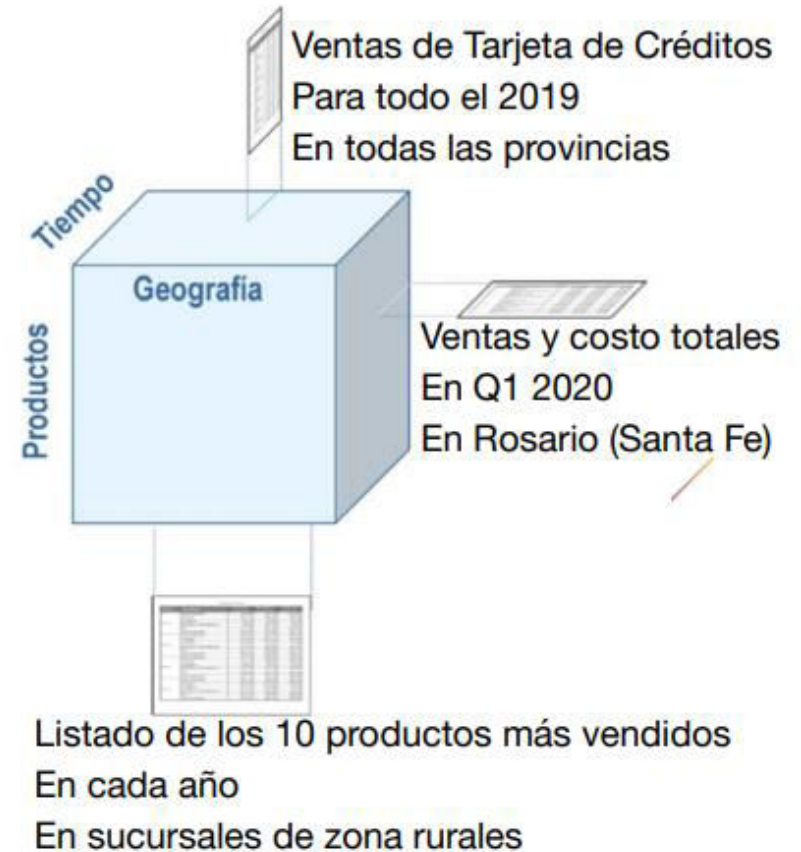
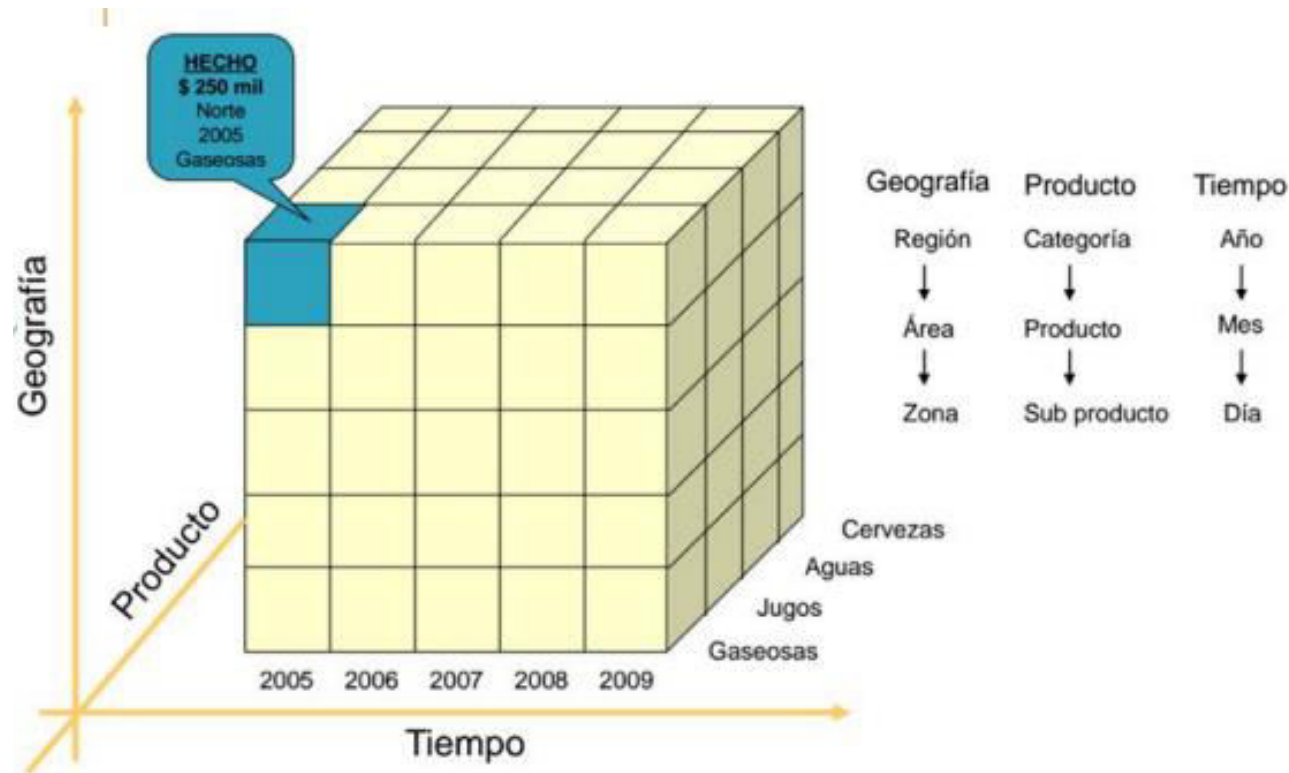
Ejercicio/Periodo	Ratios			001.2013		000.2012	
	Actividad Plan	Actividad Real	Desviación	Actividad Plan	Actividad Real	Desviación	Actividad Plan
Resultado total	877,846,989.03	1,588,697,578.56	81.0 %	269,534,095.21	471,218,831.75	62.5 %	152,987,583.63
0001/000000000	26,571,951.18	27,858,863.06	33.4 %	1,226,785.88	7,097,600.88	36.1 %	7,283,691.52
0001/000000110	26,571,951.18	27,858,863.06	33.4 %	1,226,785.90	7,097,600.88	36.1 %	7,283,691.52
0001/000000117	35,366,284.32	27,858,863.06	81.4 %		7,097,600.88		7,283,691.52
0001/000000127	3,134,429.48	200,301,394.30	6,233.7 %	1,200,786.24	51,294,675.48	871.6 %	
0001/000000301	84,040.43	462,074.32	466.8 %	940.00	288,140.38	28,392.2 %	
0001/000000302	45,377.28	462,074.32	801.8 %	460.00	238,440.64	47,311.6 %	
0001/000000304	543,166.88	1,435,113.98	264.3 %	660.00	614,140.64	460,140.64 %	
0001/000000300	164,991.87	1,961.40	11.9 %				
0001/000000301	3,275.73	-0,636.99	-19.4 %				
0001/000000302	181,490.88	107,943.17	59.5 %				
0001/000000300	626,389.16	822,767.06	131.4 %				
0001/000000301	1,643,174.96	807,351.40	49.1 %				
0001/000000300	35,433.04	35,438.71	100.0 %				
0001/000000304	82,505,504.20	34,485,381.05	41.8 %				
0001/000000300	1,609,187.72	1,403,670.00	87.2 %				
0001/000000400		2.14					

Cubo de Ventas



Niveles de realización de BI

Análisis OLAP/Multidimensional



Visualización de datos

“La **Visualización** es toda representación de datos en forma de gráfico, diagrama, matriz o tabla, un número o texto, y todos esos elementos que aplicados correctamente nos ayudarán a comprender una situación, y en conjunto nos permitirán una toma de decisiones informada y eficiente.”

“La visualización de datos es el proceso de convertir tus datos en representaciones gráficas que comunican relaciones lógicas y conducen a una toma de decisiones más informada.”

- Permite a los tomadores de decisiones clave ver análisis complejos en un diseño visual, para que puedan identificar nuevos patrones o captar conceptos desafiantes.
- Desde las métricas de la página web y el rendimiento del equipo de ventas hasta los resultados de la campaña de marketing y las tasas de adopción de productos, hay una variedad de puntos de datos que tu organización necesita rastrear.
- Cuando tienes varios proyectos a la vez, necesitas un método de informe rápido y efectivo que te permita tener claridad de lo que tienes que realizar.

Ejemplos de gráficas para un dashboard



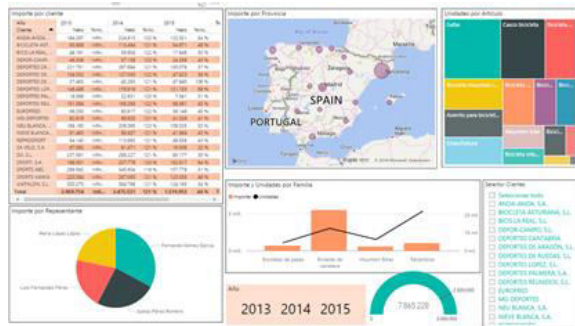
Tipos de visualización de datos

Existen multitud de **técnicas de análisis** y aproximaciones según la naturaleza del dato. Por lo general, podemos establecer tres tipos de **visualizaciones gráficas**:

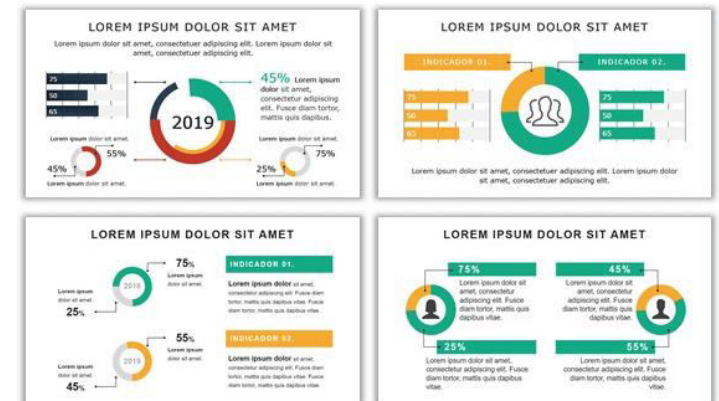
1. Elementos básicos de representación de datos. Un elemento básico de visualización puede ser un gráfico, un mapa, un KPI, tablas de datos, un grafo, etc.



2. Cuadro de mando. Es una composición compleja de visualizaciones individuales que guardan una coherencia y una relación temática entre ellas. Son ampliamente utilizados en las organizaciones para análisis de conjuntos de variables y toma de decisiones.



3. Infografías. Es una composición de visualizaciones que construye, a partir de diferentes elementos, un significado complejo para el usuario. Estas están más enfocadas a la construcción de narrativas a partir de los datos; es decir, para contar historias.



Visualización de datos

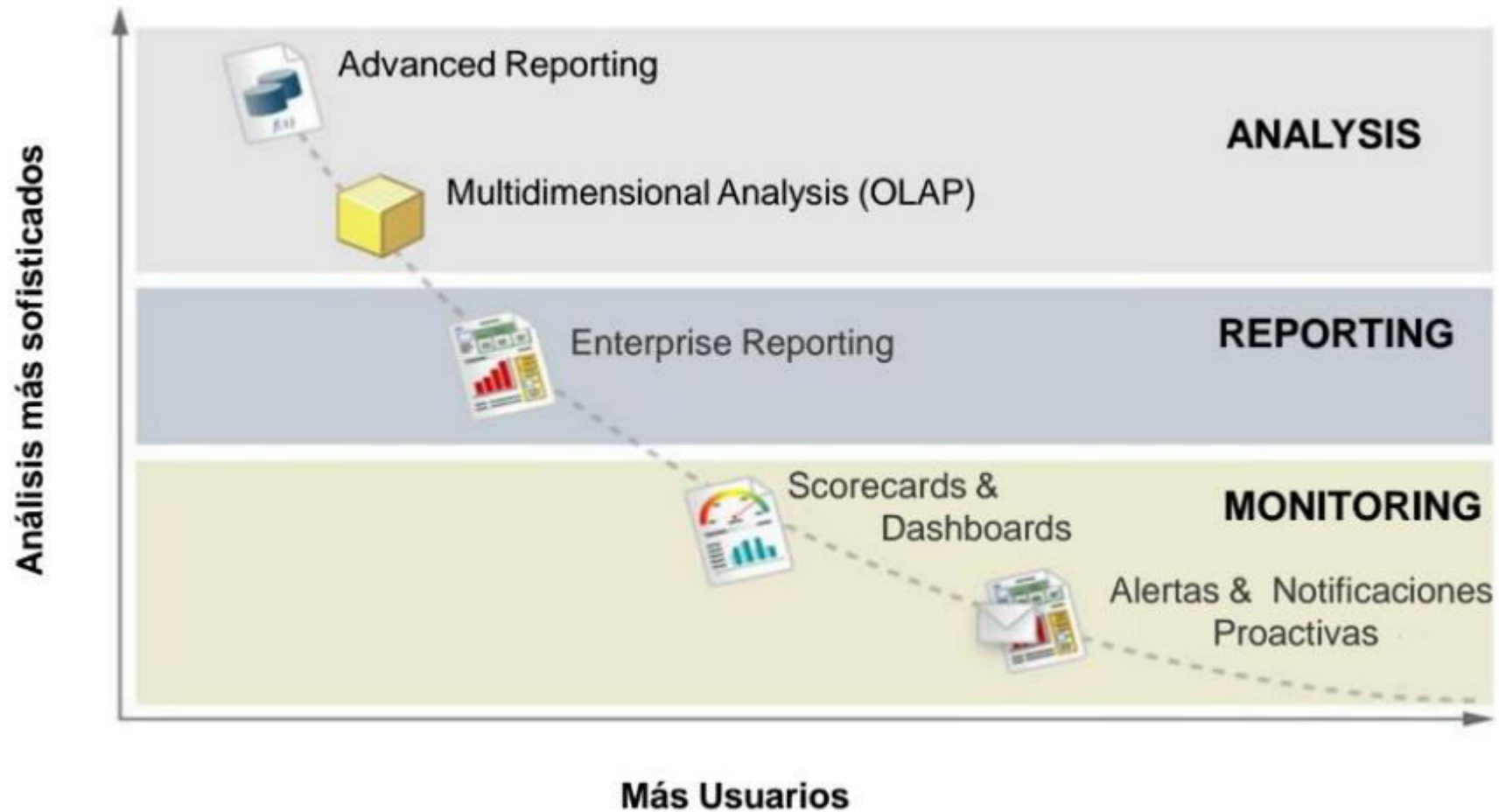
Consejos para conseguir una visualización de datos óptima

Por lo general, cuando llegue el momento de **presentar los datos extraídos** durante el proceso de análisis es recomendable seguir estas pautas:

Utilizar un gráfico adecuado. No todos comunican o expresan lo mismo y pueden confundir o no ajustarse a lo que quieres comunicar.

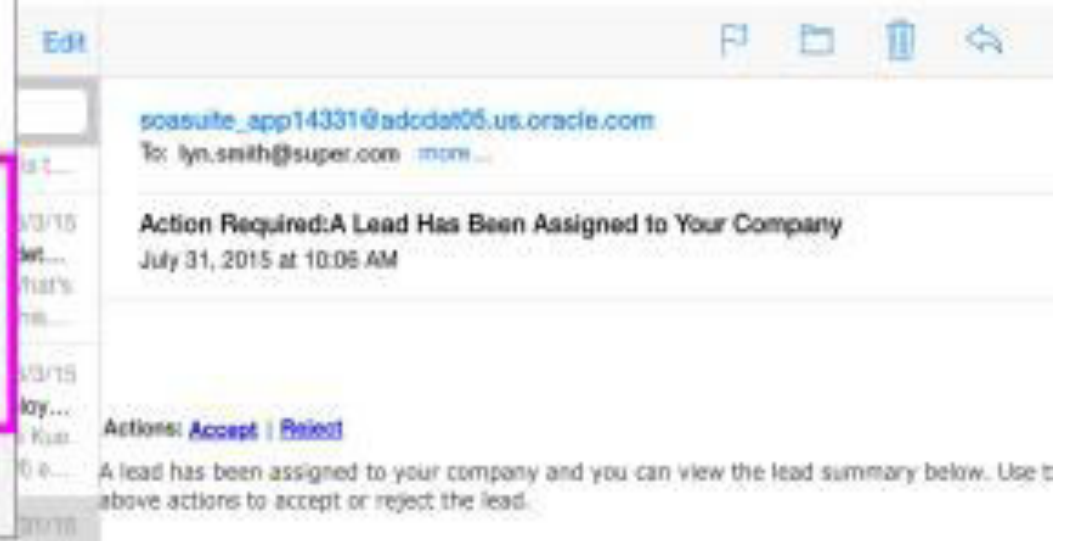
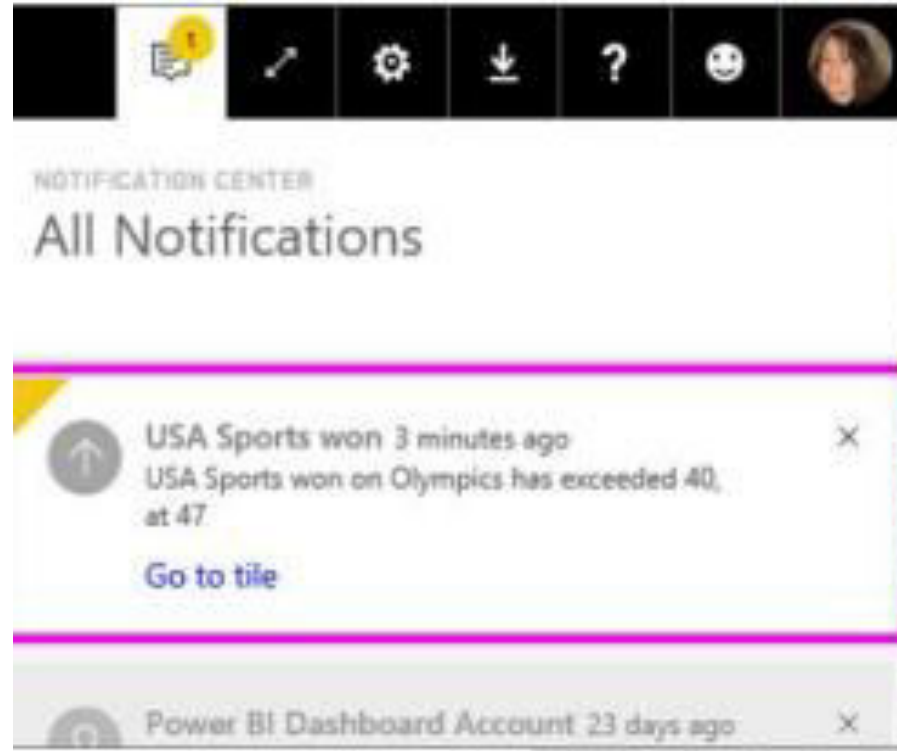
- **Resaltar la información más importante.** Ayuda al lector en el proceso de comprensión de la información y de las conclusiones.
- **Evitar la representación de los datos por medio de tablas.** En el caso de utilizarlas, asegurarse de incluir colores, tamaños y elementos distintivos que faciliten su entendimiento.
- **Centrarse en sacar conclusiones.** Al final esto es lo primordial en el proceso analítico.
- **Destacar lo positivo y lo negativo.** De ambas se extraen enseñanzas: se deben resaltar las acciones que han provocado la obtención de ratios positivos y aprender de las acciones que han desfavorecido dicho propósito.
- **Reunir a las partes implicadas en la mejora del proceso.** Así ofrecemos una visión general a todos aquellos que participan en la evolución y desarrollo de la empresa.
- **Proponer “next steps”** u objetivos de mejora para el siguiente período.
- **Hacer comparativas** tanto de meses anteriores como de años previos al actual, para identificar y medir tendencias.

Business Intelligence



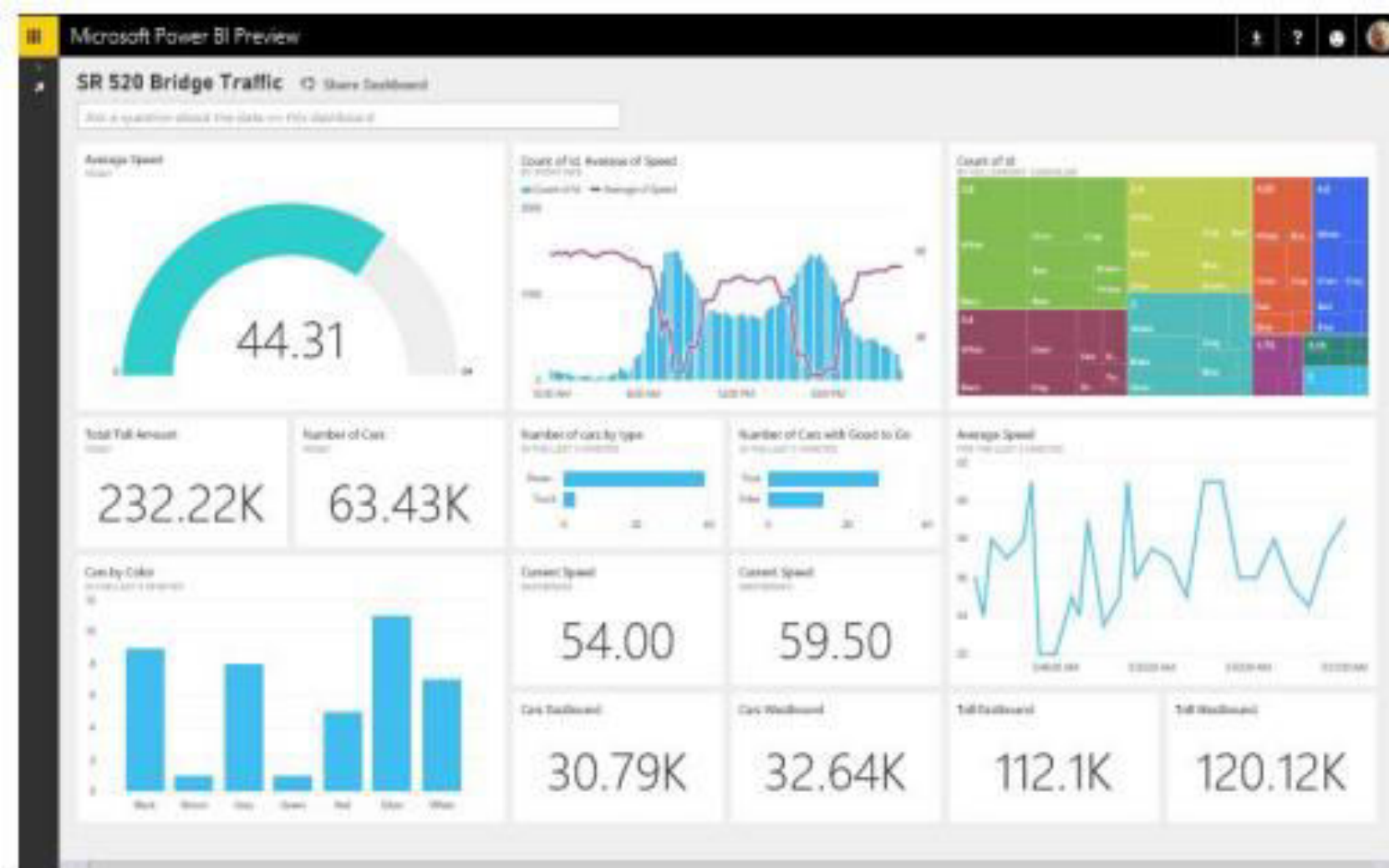
Niveles de realización de BI

Monitoreo:
Alertas &
Notificaciones



Niveles de realización de BI

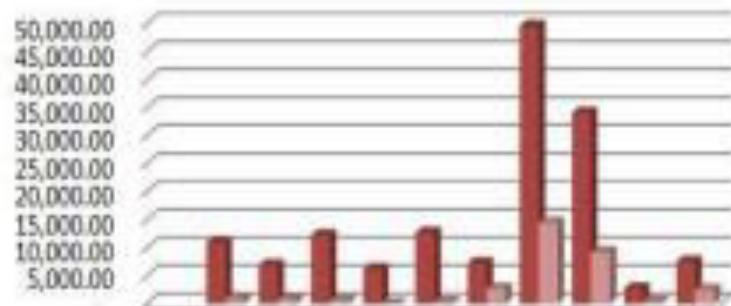
Monitoreo: Dashboards (Tableros) & Scorecards



Niveles de realización de BI

Reportes

Actual Attendance	Prior YTD					YTD					Yearly Growth
	2010 Q 1	2010 Q 2	2010 Q 3	2010 Q 4	Total	2011 Q 1	2011 Q 2	2011 Q 3	2011 Q 4	Total	
Child Admission	5,849	44,668	50,269	905	101,721	5,747	40,791	47,958	774	96,268	-8.34%
Adult Admission	12,188	78,431	65,329	1,749	177,767	12,803	75,973	83,290	1,690	173,846	-2.21%
Membership Admission	54,829	238,444	178,764	5,088	477,123	50,070	227,887	170,824	5,265	453,646	-4.92%
Group Sales Tickets	2,678	22,732	24,182	353	49,945	2,861	19,776	22,348	576	45,661	-8.66%
Group Sales - Events	744	50,484	10,367	830	22,425	1,368	9,503	12,698	465	23,664	5.53%
Comp Admission	958	17,965	9,518	185	28,644	1,727	21,088	9,177	238	32,230	12.52%
School Admissions	3,805	78,409	7,898	895	90,997	3,381	68,582	5,979	14	77,956	-14.24%
Event Admission	337	7,298	5,338		12,973	1,268	6,679	7,400	15	17,362	33.83%
Education Program Admissions	3,944	12,515	10,287	1,007	27,733	2,722	10,614	8,386	255	21,989	-20.71%
Online Tickets	1	1	125	3	130	80	2,030	427	1	2,538	1,852.31%
Total	85,131	510,997	382,117	11,113	989,358	82,247	484,323	388,297	9,293	944,160	-4.57%



Herramientas de explotación de BI



Herramientas de explotación de BI

Gartner Magic Quadrant se clasifica las empresas proveedoras de **plataformas de análisis y business intelligence (ABI)** más relevantes del sector. Los **criterios** en los que se basa Gartner a la hora de evaluar a las compañías son la **visión** y la **capacidad de ejecución**.



Gráfico cuadrangular clasifica a las empresas en 4 categorías:

- *Leaders* (líderes)
- *Visionaries* (visionarios)
- *Niche Players* (actores de nicho)
- *Challengers* (desafiadores)

As of February 2018

© Gartner, Inc

Herramientas de explotación de BI

¿Qué herramientas puedo usar para conseguir todo esto?



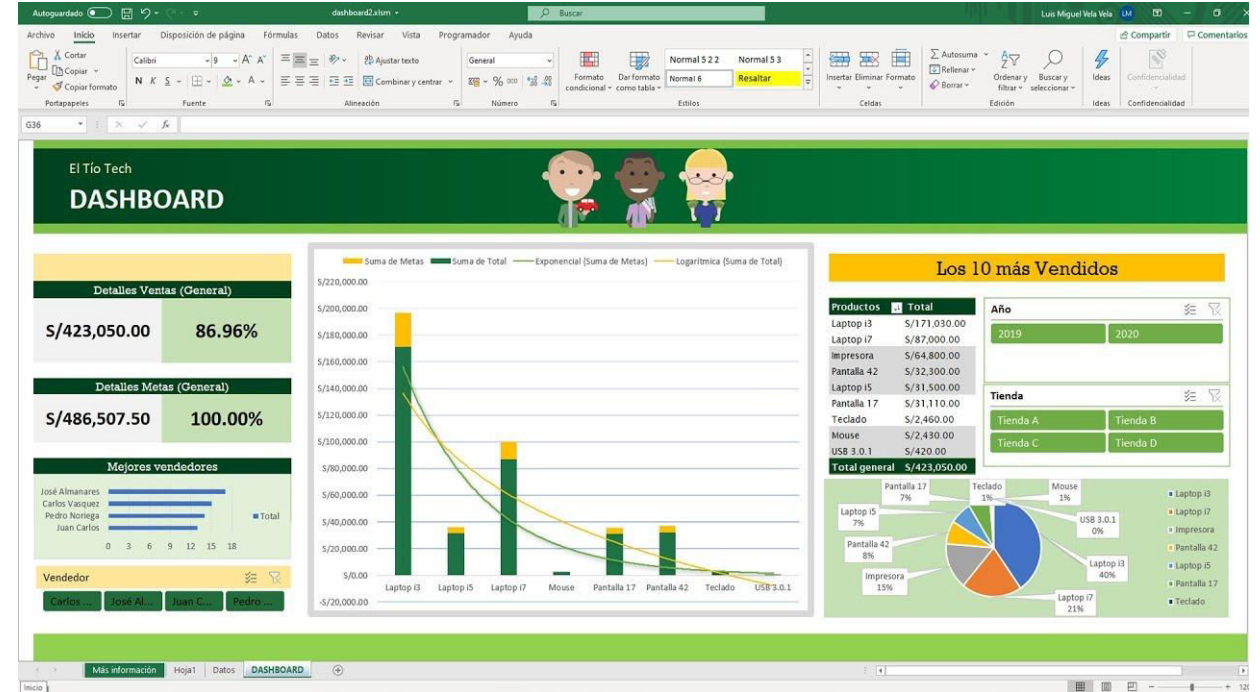
Herramientas de explotación de BI

MS Excel

Excel no es la primera herramienta en la que pensamos cuando hablamos de visualización de datos en un sistema de **Business Intelligence**, pero es la herramienta analítica más utilizada por los usuarios de negocios, ya que es muy **accesible**.

Hoy, Excel ofrece funcionalidad **end-to-end self-service BI** a través de capacidades como Power Query, Power Pivot, Power View y Power Map.

Con la accesibilidad de Excel, la rigurosidad que otorga la conexión a los servicios analíticos y la proliferación de Office 365, se reduce la barrera de entrada para las empresas que desean empezar a aprovechar los beneficios de la inteligencia empresarial.



Herramientas de explotación de BI

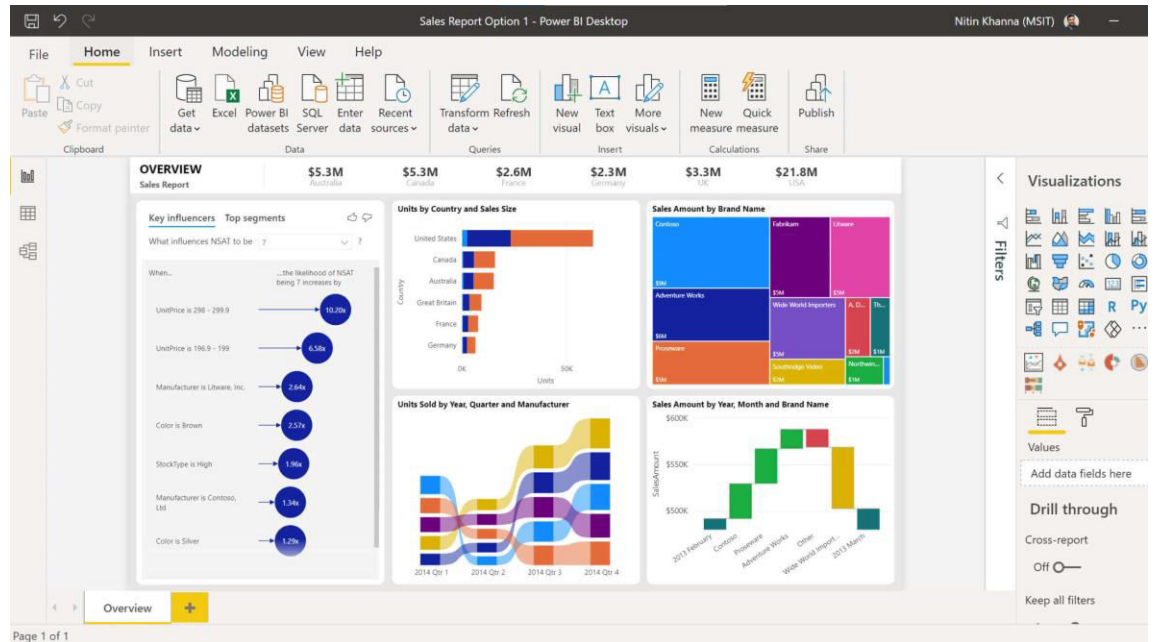
Power BI

Power BI, de Microsoft, nos permite incorporar datos de diferentes orígenes y crear un **modelo** que permita relacionarlos para su posterior **visualización**.

Tendremos la posibilidad de **analizar** nuestra información y **compartirla** con otros usuarios de la empresa tanto online como offline, en la oficina o a través de dispositivos móviles.

Ventajas:

- Integración con todo el ecosistema de Microsoft y de **Azure** (**Excel**, Teams, etc.)
- Capacidades de inteligencia artificial
- No es necesaria experiencia técnica, es muy fácil de aprender
- Accesible con aplicaciones móviles
- Automatización de reportes y limpieza de datos
- Herramienta líder en el mercado



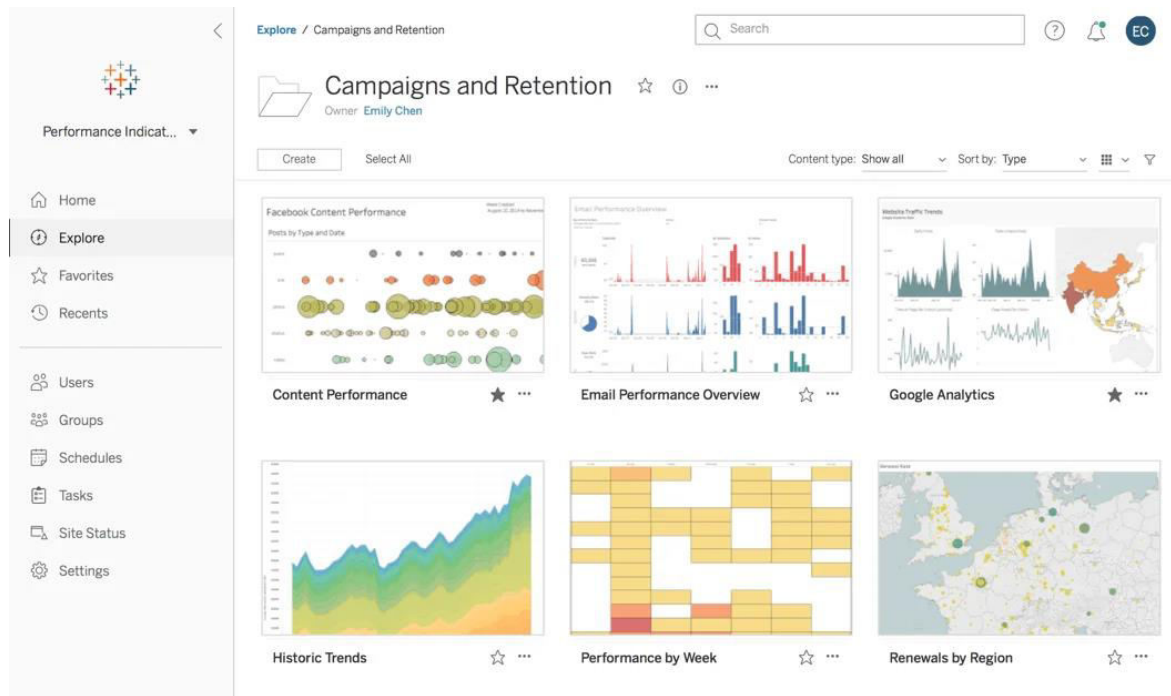
Desventajas:

- Las versiones empresariales tienen un precio elevado
- Difícil de migrar fuera del ecosistema Microsoft

Herramientas de explotación de BI

Tableau

Es una herramienta de **visualización de datos** que ayuda a **simplificar** los datos y a convertirlos a formatos muy fácilmente **comprensibles**. Esta herramienta permite la creación de **dashboards** y **worksheets**, además de un análisis de datos **rápido**. Las visualizaciones que se consiguen con Tableau son sencillas, de modo que **incluso los usuarios no técnicos** pueden desenvolverse fácilmente con la herramienta.



Ventajas:

- Integración con servicios *cloud*
- Fácil de usar (*Drag and Drop*)
- Extensible con *addons* para gobierno

Desventajas:

- Enfocado solamente en tareas de reporting, sin capacidad de **ETL y transformación de datos**.



Herramientas de explotación de BI

QlikSense/QlikView

QlikSense permite que **usuarios no técnicos** puedan visualizar datos y **hacer preguntas**. El objetivo de QlikSense es permitir **preguntar y responder** las cuestiones que surgen de la representación gráfica de los datos.

Esta herramienta nos permite integrar y combinar diversas fuentes de datos para crear informes y extraer métricas y conclusiones fácilmente.

La herramienta QlikSense es una solución basada en cloud que usa machine learning para mejorar la analítica.

Ventajas:

- Dashboards interactivos
- Despliegue en cloud
- Exportación de datos y dashboards a formatos de imagen, excel o PDF
- Acceso con API y dispositivos móviles

Desventajas:

- Curva de aprendizaje elevada para nuevos usuarios
- Soporte empresarial



Herramientas de explotación de BI

Microstrategy

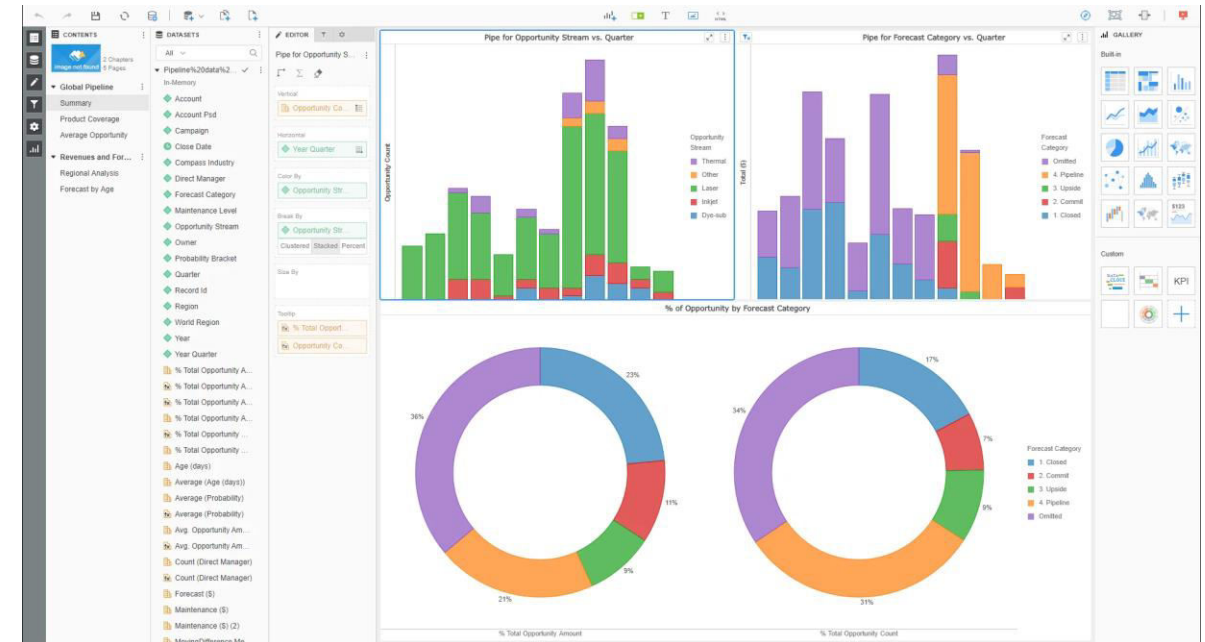
Microstrategy Workstation nos permite analizar datos visualmente sin coste de licencia

Ventajas:

- Alto grado de personalización
- Análítica de datos en tiempo real
- Microstrategy dispone de un SDK
- Minería de datos y analítica predictiva

Desventajas:

- Opciones de exportación limitadas
- Plataforma compleja de administrar



Otras herramientas interesantes a considerar al elegir una solución para BI son **Sisense**, **SAS BI**, **Pentaho BI**, **SAP BI**, **Business Objects**, **Carto**, **Domo** y **Oracle BI**.



Herramientas de explotación de BI

¿Cómo Decidir qué Herramienta de BI usar?

Muchas de estas herramientas de BI **comparten las funcionalidades básicas**. Para tomar una **buena decisión** deberemos **determinar si necesitamos capacidades adicionales**, como integración con un ecosistema en particular, por ejemplo Microsoft con Power BI o capacidades de despliegue en la nube.

Si necesitamos hacer **analítica predictiva o más avanzada** y apoyarnos en **inteligencia artificial**, deberemos valorar soluciones como Power BI o Qlik.



Herramientas de explotación de BI

¿Cómo Decidir qué Herramienta de BI usar?

Algunos de los factores más importantes para tomar esta decisión son los siguientes:

El **grado de compatibilidad técnica** con herramientas y tecnologías de almacenamiento ya usadas en la organización. Es muy importante que la herramienta de BI elegida sea capaz de conectar a bases de datos diversas directamente o a través de conectores hechos a medida.

Capacidad colaborativa: La herramienta puede ofrecer la capacidad de compartir entre sus usuarios espacios de trabajo para colaborar y compartir dashboards. También, la capacidad de definir equipos de trabajo y aislar estos espacios de trabajo entre sí.

Precio: Debemos determinar si la solución nos ofrece un precio competitivo y si aporta funcionalidades que realmente necesitamos frente a opciones gratuitas.

Facilidad de uso: Cuanto más sencilla es de usar y menos experiencia técnica requiere es más sencilla de introducir en el equipo o encontrar personas adecuadas.

BI - KPIs

Key Performance Indicators

- Ayudan a **definir y medir el progreso** hacia las metas de una organización.
- Son indicadores **clave de rendimiento**, son **mediciones**.
- Son **diferentes** para cada organización.
- Son **medibles, alcanzables, relevantes**.
- Reflejan **compromiso y responsabilidad**.

Ejemplos de KPI's



Menciones
RT's
Quotes
CTR
Respuestas y comentarios



Impresiones
Me gusta
Alcance (Reach)
Comentarios
Reacciones
Shares
Otros muros



Visitas al perfil
Impresiones
Clics al sitio web
Contacto: SMS, email, llamadas
Alcance
Stories:
Replays
Impresiones
Swipe up's
Sticker taps



Reproducciones
Suscriptores
Comentarios o "shares"
Video Integrado (embebido)
Me gusta/
No me gusta
Favoritos
Retención (tiempo de retención)



Conexiones
Vistas al perfil
Nuevas conexiones
Comentarios
Me gusta



[Video: ¿Qué es un KPI?](#)

BI - KPIs Key Performance Indicators

CARACTERÍSTICAS DE LOS KPIS EN REDES SOCIALES (SMART)



BI - KPIs Key Performance Indicators

¿Cuál es la diferencia entre KPI y métrica?

Esta es una confusión muy común.

KPIs no son lo mismo que métricas, pero una métrica puede convertirse en un KPI.

Los **KPIs** son indicadores importantes para tu negocio y su objetivo, una **métrica** es apenas algo para ser medido.

Si por algún motivo esa métrica se vuelve relevante para tu estrategia, se convierte en un indicador clave.

Lo importante es entender qué es lo que puede ayudar en la toma de decisiones dentro de tu empresa. **Esta es la premisa básica para escoger cualquier KPI y es así como una métrica se convierte en un indicador.**



TP N° 7: Ejemplos de visualizaciones



-Buscar ejemplos de visualizaciones (gráficos más convenientes) con las que trabajen o hayan trabajado. Armar una presentación de no más de 3 minutos por alumno o por grupo (a definir según los participantes).

Armar una presentación en 1 hoja en base a lo solicitado para exponer en clase.

Éxitos!!