

## **UNIDAD 3: BUSINESS INTELLIGENCE Y LA SOCIEDAD DE LA INFORMACIÓN**

**Módulo Profesional:**  
**Análisis de grandes datos (Big Data)**

## ÍNDICE

RESUMEN INTRODUCTORIO.....	4
INTRODUCCIÓN.....	5
CASO INTRODUCTORIO .....	6
1. INTRODUCCIÓN AL SISTEMA DE BI .....	8
1.1 El nacimiento de la sociedad de la información.....	8
1.1.1 Indicadores y políticas para la sociedad de la información .....	18
1.1.2 La sociedad de la información en España .....	27
1.2 Regulación en materia de protección de datos .....	31
1.2.1 GDPR.....	32
1.2.2 GDPR y big data .....	38
1.3 Los recursos humanos en la sociedad de la información en España..	43
1.4 Los sistemas de BI .....	54
1.4.1 Inteligencia de negocios en la era de la transformación digital...	56
1.4.2 Cómo entender el business intelligence en tiempos de big data .	57
1.4.3 ¿Cuándo se requiere la inteligencia de negocios BI? .....	59
1.4.4 Como definir la estrategia de business intelligence .....	60
1.4.5 ¿Cómo saber en qué nivel de la inteligencia de negocio se encuentra mi organización?.....	61
1.4.6 Retos que afrontar.....	66
2. ARQUITECTURA DE SISTEMAS DE BI .....	68
2.1 Antecedentes .....	69
2.2 Ventajas del uso de BI .....	75
2.3 El análisis de datos en BI .....	76
2.3.1 La obtención de la información.....	77
2.3.2 Gestión de la información.....	82
2.3.3 Distribución de la información .....	85
2.4 Data warehouse.....	87
2.4.1 Profundizando en la arquitectura del data warehouse .....	90
2.4.2 Variaciones en la arquitectura de data warehouse .....	95
2.5 Data mining (minería de datos) .....	98
2.6 OLAP.....	101
2.7 EIS .....	103

2.8 El sistema de reporting .....	103
2.9 Sistemas de apoyo a la decisión (DSS) .....	105
2.9.1 Fases de un proceso de toma de decisiones .....	107
2.9.2 Tipos de DSS .....	109
3. BUSINESS ANALYTICS: CLUSTERING.....	111
3.1 Tipos de clustering .....	113
3.2 Tipos de algoritmos de clustering.....	114
3.2.1 Método K Clustering .....	115
3.2.2 Método jerárquico de agrupación / Hierarchical clustering .....	115
3.2.3 Diferencias entre el método de clustering K y el jerárquico .....	116
3.2.4 Aplicaciones comunes de los algoritmos de agrupación .....	117
3.2.5 Ejemplo: Clusterizar ingresos de hogares mediante K clusteres.	
.....	117
3.2.6 Clustering con Power BI.....	121
4. BUSINESS ANALYTICS: CLASIFICACIÓN.....	122
4.1 Introducción a Power BI .....	122
4.2 Componentes de Power BI Los componentes de Power BI son: .....	124
4.3 Primeros pasos con Power BI Desktop .....	125
4.4 Conectar con archivos de datos y transformar los datos .....	130
4.5 Transformar las consultas en Power BI .....	151
4.6 El lenguaje DAX en Power BI.....	163
5. TENDENCIAS EN BUSINESS INTELLIGENCE .....	167
5.1 RPA .....	169
5.2 Zapier .....	171
RESUMEN FINAL .....	177

## RESUMEN INTRODUCTORIO

En esta unidad, basada en el estudio de la sociedad de la información y el business intelligence, comenzaremos analizando cómo ha sido la evolución de la sociedad, particularmente en España, pero sin olvidar el ámbito de influencia en el cual vivimos, que es la Unión Europea.

Comprenderemos las dinámicas y factores que han provocado la actual importancia de los datos y de la información en todos los ámbitos.

También, estudiaremos aspectos éticos y legales en materia de manejo y gestión de los datos. Sin duda, una sociedad moderna y de derecho está obligada a considerar con toda relevancia cuestiones como la privacidad y el derecho de las personas a decidir quién, cómo y cuándo puede acceder a sus datos personales, por lo que veremos el reglamento general de protección de datos, que otros países y grandes bloques comerciales como Estados Unidos están imitando.

Además, conoceremos la realidad de los recursos humanos en la sociedad de la información, ya que gran parte del éxito empresarial dependerá de las personas de las que se rodee y cómo lo gestione.

Acto seguido estudiaremos los sistemas de business intelligence, tanto desde un punto de vista funcional y de impacto en el negocio, como desde la perspectiva de su arquitectura y sus tendencias.

Este recorrido por la arquitectura de los sistemas de información nos ayudará a tener una visión general de cómo se integran las diferentes piezas que conforman una solución de inteligencia de negocio, especialmente en lo referido al data warehouse, que es la piedra angular de este tipo de soluciones.

Aunque posteriormente, tendremos ocasión de profundizar con mucho más detalle en los aspectos técnicos, esta unidad es el lugar adecuado para adquirir una primera visión holística de las plataformas de BI.

Cómo complemento a los repositorios o almacenes de datos, presentaremos los restantes módulos necesarios para configurar una plataforma completa de inteligencia empresarial, como son las herramientas de minería de datos, los sistemas de generación de informes, las herramientas de soporte a la toma de decisiones en general y, específicamente, aquellas que están orientadas a prestar soporte a los niveles más altos de toma de decisión empresarial, que son los EIS.

Tras ello, conoceremos dos herramientas para análisis de los datos empresariales: la clasificación y la clusterización. Estos tipos de algoritmos se utilizan desde hace décadas, pero han ganado mucha importancia, entre otros motivos, por la democratización del machine learning actual.

Finalmente, tendremos una toma de contacto con una tecnología que, sin ser exclusivamente una herramienta de BI, sí que complementa y habilita a bajo coste unas capacidades de integración de datos avanzadas que hasta hace poco tiempo eran imposibles o tremadamente costosas. Nos estamos refiriendo a los robots software, o RPA, para la automatización de procesos como pueden ser la depuración de datos, la carga de los mismos, el acceso a fuentes remotas o la agilización de los sistemas heredados (legacy) etc.

## INTRODUCCIÓN

La segunda mitad del siglo XX y, especialmente, sus dos últimas décadas han visto suceder un drástico cambio social, donde la información y el conocimiento se ha convertido en el activo más valioso de las personas y de las organizaciones.

A este auge de la sociedad de la información han contribuido tanto los gobiernos, como las empresas y las instituciones científicas y académicas. En paralelo a la eclosión de la información se ha creado una nueva rama de la tecnología denominada business intelligence o inteligencia xde negocio. Esta disciplina, que combina capacidades informáticas y analíticas, ha permitido el surgimiento de nuevos negocios, así como la transformación de empresas consolidadas.

Las arquitecturas de sistemas que dan soporte y proporcionan la infraestructura necesaria para llevar a cabo la inteligencia del negocio se sustentan sobre una pieza angular que son grandes repositorios de información histórica denominados data warehouse.

En la actualidad es muy difícil encontrar empresas con una relevancia mínima que no hayan desarrollado una estrategia más o menos ambiciosa en materia de business intelligence.

A pesar de esa consolidación en el mercado tecnológico, la realidad es que el futuro del business intelligence es apasionante por la cantidad de retos que debe afrontar y por las nuevas capacidades que la tecnología nos ofrece, como son la inteligencia artificial, la nube, los nuevos modelos de computación, etc. y ese horizonte, todavía de futuro, pero cada vez más cercano, que es la computación cuántica y que todo lo transformará.

En el ámbito del business intelligence y, más concretamente, en lo que tiene que ver con la minería de datos, dos son las herramientas más divulgadas y utilizadas.

Por un lado, el clustering o agrupación automática de elementos, que nos permite explorar conjuntos de datos y encontrar en muchas ocasiones conocimiento oculto.

Por otro lado, los algoritmos y herramientas tecnológicas para la clasificación automática, que permiten acelerar procesos, suplir al humano e, incluso, mejorar la precisión de la clasificación realizada por personas por motivos como, por ejemplo, la eliminación de los sesgos de pensamiento propios de la subjetividad humana.

## CASO INTRODUCTORIO

Trabajas en el departamento de tecnología de una empresa de venta de muebles que en los últimos años está viendo como las ventas se reducen. Los compradores más jóvenes no se interesan por sus productos y su posicionamiento y valor de marca está decreciendo.

Ante esa situación, el director comercial de la empresa ha recibido una propuesta por parte de un proveedor de servicios de marketing que le permitiría adquirir una base de datos de leads (contactos comerciales potenciales) sobre los que podría dirigir sus esfuerzos de comunicación para intentar convertirlos en clientes.

El director comercial ha visto en ti a la persona adecuada para llevar a cabo este proyecto, te conoce y considera que eres la persona más innovadora del departamento.

Aceptas el reto y solicitas tiempo para analizar la situación y presentar un plan de trabajo con todos los elementos a tener en cuenta para mejorar el éxito comercial de la empresa y, particularmente, aprovechar del modo más eficiente posible la base de datos de leads que han ofrecido.

Hasta el momento tu experiencia profesional había sido como informático, sin tener que preocuparte nunca de cuestiones más abstractas. Por ello, decides estudiar más sobre la sociedad de la información y cómo todos esos cambios están afectando a los negocios en general.

Al finalizar esta unidad conocerás los aspectos necesarios a tener en cuenta para:

- Interpretar la influencia de los aspectos sociales y humanos a la hora de desarrollar un proyecto tecnológico de inteligencia de negocio.
- Identificar los aspectos legales, especialmente en materia de protección de datos, que pueden limitar un proyecto de business intelligence.
- Determinar cuál es el nivel de madurez que una organización determinada tiene en materia de inteligencia empresarial.
- Hacer un inventario de las fuentes de datos e información disponibles, así como estimar las ausentes en la empresa.
- Plantear la carga de datos ordenada en un sistema de BI.
- Diferenciar las principales arquitecturas de un data warehouse.
- Distinguir diferentes DSS, EIS y sistemas de reporting.
- Saber cuándo y por qué aplicar un algoritmo de clasificación o uno de clusterización e interpretar los resultados obtenidos.
- Manejar la funcionalidad básica de la herramienta Microsoft Power BI.
- Utilizar una tecnología de RPA y programar una primera automatización de procesos.

# 1. INTRODUCCIÓN AL SISTEMA DE BI

*En el proyecto de la empresa donde trabajas te preguntas si existe algún factor de brecha digital entre tu clientela. ¿Cuáles son los grupos de edad más frecuentes? ¿Crees que cuentan con las habilidades digitales suficientes para ser autónomos en el uso de las nuevas tecnologías? ¿Consideras que se encuentran cómodos y seguros navegando por vuestra página web o comercio electrónico?*

Según Michael Porter, el profesor de la Universidad de Harvard de quien se habla en otras ocasiones, el contexto de la sociedad de la información ha obligado a construir nuevos métodos para explotar los datos de una empresa en particular, o de cualquier tipo de organización en general, y conseguir, además, que dichos datos potencien toda la cadena de valor de la compañía. El profesor Porter no duda de que la inteligencia de negocios es una buena herramienta, aunque probablemente no la única, para responder a las necesidades identificadas.

En cierto modo podemos ver al business intelligence como la evolución de los primeros sistemas de soporte a la toma de decisiones, que no eran capaces ya de afrontar los retos que la sociedad de la información les imponía por haberse quedado obsoletos.

Esta perspectiva de Michael Porter es totalmente coherente con la de Howard Dresden, de Gartner Group, quien "bautizó" al business intelligence con el nombre propio que ha llegado hasta nuestros días, y que ya en 1989 definió a la nueva disciplina del business intelligence como el conjunto de elementos para mejorar la toma de decisiones empresarial mediante sistemas de información digitales.

## 1.1 El nacimiento de la sociedad de la información

Para todo profesional de las tecnologías de la información es importante conocer los aspectos teóricos que definen la sociedad donde se desarrolla su actividad, así como la evolución histórica que nos ha traído hasta el momento presente.

El origen de la explosión de la información, de la cual se deriva el posterior concepto de la sociedad de la información, empieza ya en el siglo XVII, con la invención de la revista científica. En ese momento, las grandes academias de la ciencia empezaron a considerar que la ciencia sólo puede avanzar mediante un intercambio de ideas abierto y transparente, respaldado por pruebas experimentales, de manera que se fomentó la publicación y difusión de las teorías científicas.

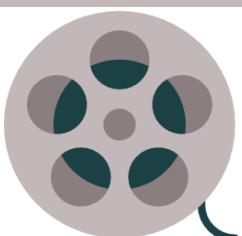


### ENLACE DE INTERÉS

Si quieras conocer más sobre el funcionamiento de las revistas científicas puedes hacerlo en el siguiente programa de Radio Uned.

<https://www.youtube.com/watch?v=KY-jKXgy02w>

A finales del siglo XIX, se producen cambios en los procesos de reproducción de documentos, que contribuyen a la edición y distribución a gran escala de los mismos y a la aparición de la cultura de masas.



### VIDEO DE INTERÉS

En el siguiente vídeo se explica qué es la cultura de masas:

<https://www.youtube.com/watch?v=4RSB0OazcFs>

Pero, sin duda, el punto de inflexión en lo que se refiere a la producción y gestión de información lo encontramos en la Segunda Guerra Mundial:

- Se intercambiaron más de 60.000 cartas e informes entre el Reino Unido y Canadá.
- El Air Documents Research Center de Londres recopiló cada día, durante la guerra, entre 4.000 y 5.000 documentos procedentes de la documentación requisada al enemigo.
- Al finalizar la guerra, el archivo nacional del Reino Unido contaba con 300.000 documentos japoneses y 13 millones de páginas interceptadas.

Todo esto llevó a que en los años cincuenta comenzasen a desarrollarse los primeros estudios sobre el crecimiento de la producción científica y, en los sesenta, se consolida la importancia de la información como un instrumento para el desarrollo económico y social.

La sociedad de la información comenzó a desarrollarse en los años 60, debido al incremento de la información y a la necesidad del desarrollo social en base al uso de las tecnologías de la información.

Aunque se han realizado numerosas y variadas definiciones de la sociedad de la información, podemos diferenciar la sociedad de la información a través de los elementos que la caracterizan:

- La tecnología: Algunos teóricos sobre la materia, al hablar sobre la sociedad de la información, prestan especial atención a los aspectos relacionados con la informática y las comunicaciones. Las tecnologías de la información y la comunicación (TIC) son, sin duda, la base de la sociedad de la información porque:
  - La tecnología facilita el almacenamiento y procesamiento de la información de forma barata.
  - Las redes de comunicaciones facilitan la transmisión de la información.
- La economía de la información: En los años 60, Fritz Machlup, en su obra *La Producción y Distribución del Conocimiento en EE.UU.*, estableció las medidas para evaluar el tamaño y la amplitud de las industrias de la información. Machlup definió las industrias de la información de forma estadística de acuerdo con cinco ramas:
  - Educación y bibliotecas.
  - Medios de comunicación.
  - Equipos informáticos.
  - Servicios de información.
  - Otras actividades de información.

Trabajando con estas categorías descubrió que cada una tenía un valor económico que repercutía en el Producto Interno Bruto (PIB), llegando incluso a demostrar que el 29% del PIB de los Estados Unidos procedía de estas industrias. Esto lo llevó a concluir que el conocimiento es una pieza fundamental de la economía moderna y que se encuentra íntimamente vinculado con otros elementos como:

- El empleo: Un aspecto que mide la aparición de la sociedad de la información son los cambios en el tipo de actividades laborales y profesionales. Porque en la sociedad actual el empleo predomina en el sector de la información.
- El tiempo-espacio: Las redes de información conectan a las personas sin importar tiempo y el espacio. Gracias a internet, la fibra óptica, los satélites, etc. las comunicaciones se han vuelto instantáneas y los repositorios de información son accesibles desde cualquier lugar.

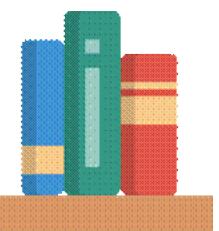
- La cultura: El enfoque cultural en una sociedad de la información resulta fácil de identificar. Se puede afirmar que la cultura contemporánea es más informativa que sus predecesoras.



#### BIBLIOGRAFÍA RECOMENDADA

Machlup, F. (1962). *La producción y distribución del conocimiento en los Estados Unidos*. Princeton University Press.

Realmente, aunque el impacto de la información y el conocimiento ya se estudiaba en los años 60, no es hasta los años 80 cuando se utiliza por primera vez la expresión “sociedad de la información”, en el ya clásico libro de Masuda, *The Information Society as Post-Industrial Society*.



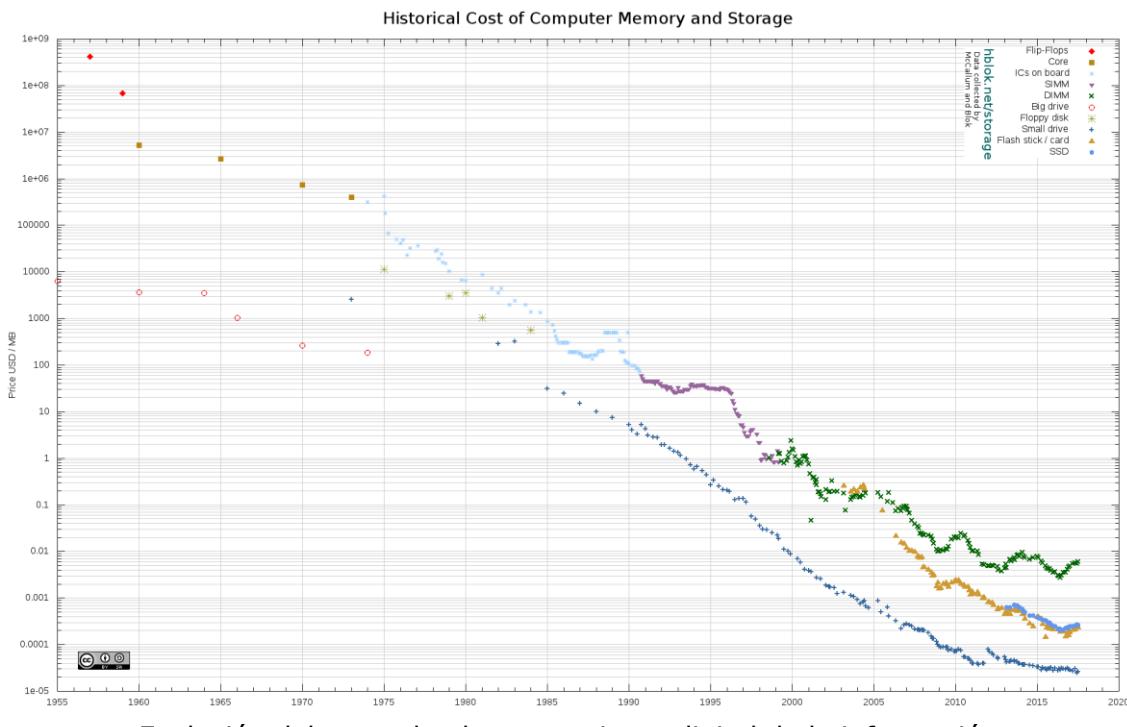
#### BIBLIOGRAFÍA RECOMENDADA

Masuda, Y. (1980) *La Sociedad de la Información como Sociedad Post Industrial*. Madrid. Fundesco-Tecnos.

El empresario y político Joan Majó, ve en la sociedad de la información un nuevo modelo social, en el cual “las consecuencias de la revolución de las tecnologías de la información y de las comunicaciones van más allá del ámbito puramente tecnoeconómico: es un proceso que implica un cambio de modelo social”.

Además, Majó ve la sociedad de la información como el elemento clave para comprender la sociedad actual en el ámbito socioeconómico. Para Majó, las principales características de este modelo de sociedad son las siguientes:

- El coste de los elementos materiales necesarios para procesar, almacenar y difundir información tiende a cero.



Evolución del coste de almacenamiento digital de la información

Fuente: hblock.net

- Utilización del código digital para almacenar o transmitir todo tipo de información.
- La liberalización de las telecomunicaciones: de un servicio público se ha pasado a un modelo basado totalmente en el mercado.
- La introducción masiva de la tecnología en el proceso de producción nos permite ahorrar recursos materiales y mano de obra.
- Se consumen menos productos y se utilizan más servicios.
- El desarrollo de una conciencia global de que los recursos materiales no son ilimitados y que la capacidad de la naturaleza para reabsorber los residuos se agota.
- La incorporación de nuevos conocimientos a lo largo de toda la vida es una fuente de progreso.
- Surge un nuevo modelo de trabajo más flexible y personalizado.
- Crecimiento de la riqueza global y de las bolsas de pobreza.

Por lo tanto, en este momento nos encontramos inmersos en una sociedad en la cual una parte significativa de la actividad humana –y de la generación de la riqueza- proviene no solo de la producción y manipulación de la información, sino también de su uso adecuado.

Formamos parte, por tanto, de una sociedad que crece y prospera a partir del uso de la información y de las tecnologías asociadas.

De forma resumida, podemos definir la sociedad de la información como aquella en la que:

- La información es un recurso económico.
- Aumenta del uso de la información por parte del gran público.
- Se desarrolla un nuevo sector de la información (tecnología y contenidos).

Pero que, además, tiene como rasgos principales:

- Se desarrolla una nueva visión del tiempo y del espacio.
- Los procesos económicos se renuevan.
- Surge una nueva interpretación geográfica y política de la sociedad.

Con frecuencia oímos hablar de sociedad postindustrial, sociedad red, sociedad de la información, sociedad del conocimiento, era de la información... pero ¿son expresiones sinónimas? ¿Hacen referencia a los mismos conceptos?

La realidad es que, aunque algunos de ellos son bastante similares entre sí, entre todos hay matices que los diferencian:

- Sociedad postindustrial es un concepto básicamente sociológico y económico, que describe un sistema social y económico que ha superado los cánones de la revolución industrial y donde los servicios superan en importancia a la producción industrial y se ha iniciado un proceso de globalización y liberalización de los mercados.
- La sociedad red recalca la importancia del movimiento de la información a través de redes, especialmente gracias a internet y todas las consecuencias derivadas de ello.
- La sociedad de la información es cuando las tecnologías para producir y distribuir información son preponderantes e influyen en todos los órdenes de la sociedad, desde la cultura a la economía. De todos los mencionados, es el concepto que más popularidad ha conseguido.
- La sociedad del conocimiento pone el acento en los individuos que han desarrollado capacidades para discernir entre toda la información disponible y que saben utilizarla en su beneficio económico, profesional, social, etc.
- Era de la información es el período cronológico que va desde la era espacial hasta la economía del conocimiento y se basa en la gran difusión que las TIC tuvieron. Fue acuñado por el sociólogo Manuel Castells.

**CITA**

“

“Al hablar de sociedad del conocimiento nos estamos refiriendo a la constitución de un nuevo paradigma tecnológico.

Dicho paradigma tiene dos expresiones tecnológicas concretas y fundamentales: una es internet. Internet no es una energía más; es realmente el equivalente a lo que fue primeramente la máquina de vapor y luego el motor eléctrico en el conjunto de la revolución industrial.

La otra es la capacidad de ingeniería genética, el concomitante ADN o la capacidad de recodificar los códigos de la materia viva y, por tanto, ser capaz de procesar y manipular la vida. (...)

Las dos revoluciones se fusionan e interactúan y se convierten en una revolución en la que todos los procesos de la información, incluso los códigos de la materia prima pueden ser programados, desprogramados y reprogramados de otra forma.”

Manuel Castells. La dimensión cultural de Internet.

**Sociedad postindustrial****Sociedad red****Sociedad de la información****Sociedad del conocimiento**

Evolución social

Fuente: Elaboración propia

Evidentemente, la sociedad de la información supone muchas ventajas a todos los niveles: facilita que se trabaje mejor, que se vive mejor, se promueve el derecho a la información, así como el derecho a la libertad de expresión... pero también plantea una serie de problemas o limitaciones: la dificultad para determinar los límites de poder, la protección de los datos personales, riesgos de exclusión, etc.

En este momento, el problema más significativo al que nos tenemos que enfrentar como consecuencia de la sociedad de la información es la brecha digital. Pero ¿en qué consiste este problema?

La expresión digital divide (brecha digital), apareció por primera vez en Estados Unidos a mediados de los 90. Inicialmente para aludir a la posesión y acceso (o no) a la tecnología. Con el tiempo, esta expresión fue abarcando más aspectos como veremos a continuación.



### ENLACE DE INTERÉS

Seguramente conoces las charlas TED que se organizan por todo el mundo.

En este caso, Aleph Molinari, como activista para la reducción de la brecha digital nos da su visión experta.

[https://www.ted.com/talks/aleph\\_molinari\\_let\\_s\\_bridge\\_the\\_digital\\_divide/transcript?language=es](https://www.ted.com/talks/aleph_molinari_let_s_bridge_the_digital_divide/transcript?language=es)

Así, el Departamento de Comercio de Estados Unidos la definió de la siguiente manera: "Algunas personas disponen de los ordenadores más potentes, el mejor servicio telefónico y el mejor acceso a internet, así como la riqueza de contenidos y una educación para el aprendizaje a lo largo de su vida... en cambio, hay otro grupo que no tiene acceso. La diferencia entre estos dos grupos es la brecha digital".

Otra definición significativa de esta expresión es la propuesta por la OCDE, que considera que la brecha digital "se refiere al desfase o división entre individuos, hogares, áreas económicas y geográficas con diferentes niveles socioeconómicos en relación, tanto en sus oportunidades de acceso a las tecnologías de la información y comunicación, como al uso de internet".

El hecho de que las personas puedan superar la brecha digital implica:

- Disponer de un ordenador u otro dispositivo de acceso para poder conectarse a internet.
- Poder acceder a internet desde casa y/o el trabajo.
- Conocer las herramientas básicas para poder acceder y moverse por la red.
- La capacidad de convertir la información accesible a través de Internet en conocimiento.

Por lo tanto, debemos destacar que la brecha digital va más allá del acceso o no a la tecnología, por lo que para superarla es necesaria la adquisición de una serie de habilidades o capacidades informacionales, que nos permitan obtener el máximo provecho de la sociedad de la información.

Estas habilidades hacen referencia a aspectos como:

- Saber navegar por las fuentes de información.
- Saber utilizar los sistemas de información.
- Saber discriminar la calidad de las fuentes.
- Saber discriminar la fiabilidad de las fuentes.
- Saber dominar la sobrecarga de información.
- Saber aplicar la información a problemas reales.
- Saber comunicar la información.

Es decir, con el desarrollo de la tecnología y la generalización de su uso, surge una nueva cultura informacional que requiere unas habilidades específicas. Estas habilidades se conocen con el término inglés literacy skills, e implican una nueva forma de alfabetizar a los ciudadanos. Por lo tanto, hablar de alfabetización en la sociedad de la información supone hablar de tres vertientes diferentes:

- La alfabetización tradicional, relacionada con las habilidades mecánicas de lectoescritura.
- La alfabetización funcional, que implica la capacidad de comprender y desenvolverse en el mundo laboral y personal usando las herramientas de lectoescritura.
- La alfabetización informacional, que incluye las habilidades de búsqueda y localización de información, así como las habilidades, conocimientos, destrezas y aptitudes en el uso y aprovechamiento de las tecnologías de la información, y sus contenidos, para desarrollarse en el mundo laboral y personal.

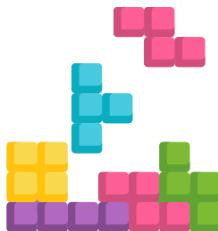
#### CITA

“

“Alfabetización informacional es saber cuándo y por qué necesitas información, dónde encontrarla, y cómo evaluarla, utilizarla y comunicarla de manera ética”.

Chartered Institute of Library and Information Professionals (CILIP)

Llegados a este punto, debemos insistir en la diferencia que existe entre las habilidades relacionadas con el uso de las tecnologías de la información y la alfabetización informacional.



### EJEMPLO PRÁCTICO

Estás destinado por tu empresa para liderar un proyecto en Bolivia, donde la principal cadena de grandes almacenes del país os ha contratado para desarrollar su plataforma de comercio electrónico. Antes de trasladarte, un compañero de oficina que trabajó allá en un proyecto previo, te ha advertido de las dificultades derivadas de la brecha digital y que pueden poner en riesgo tu proyecto.

¿Qué consideraciones deberías hacer al respecto?

### SOLUCIÓN

En primer lugar, debería analizar cuál es la disponibilidad de dispositivos para acceder a internet que hay en el país, pero no solamente quedarme con los datos medios, sino que debería realizar un análisis más detallado y conocer la realidad por áreas geográficas, segmentos de población, etc. Porque se trata de un país muy heterogéneo.

En base a esos datos, sabré si debemos orientar la plataforma de compras más a un uso desde el ordenador o el móvil.

También es necesario conocer la disponibilidad de acceso a internet y la calidad de esta, porque si es muy limitada deberemos utilizar tecnologías que, aunque menos potentes, sean asumibles por los usuarios, descartando el uso de webs con videos o animaciones.

Igualmente deberá conocer el grado de alfabetización de la población, porque probablemente, un gran número potencial de clientes de los grandes almacenes tengan dificultades de lectura, así que deberemos valorar el realizar un sistema de comercio electrónico muy iconográfico, de modo tal que incluso, personas que no saben leer, puedan realizar compras en base al uso de botones y menús que se expliquen solos.

Sin duda, la alfabetización informacional en este momento es más necesaria que nunca, debido al aumento de recursos y posibilidades de acceso a la información y es un requisito para el aprendizaje a lo largo de toda la vida.

### **1.1.1 Indicadores y políticas para la sociedad de la información**

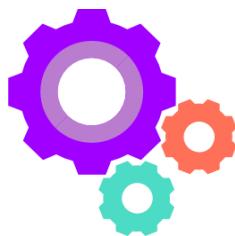
Aparte de España, ¿Hay algún otro mercado que sea importante para la empresa de muebles donde trabajas? Si es así, ¿tienes o podrías tener acceso a indicadores sobre la sociedad de la información en esos países? ¿Eres consciente de cómo puede afectar a la relación que clientes y potenciales clientes desarrollan cara a tu empresa? Tal vez una web de comercio electrónico no es la más adecuada en países donde la velocidad media de conexión a la red es baja.

¿Qué tal está funcionando la web en comercio electrónico a través de teléfonos móviles? ¿Cuál es el grado de cobertura de la red y el porcentaje de uso de smartphones en tus principales mercados de destino? Todas estas preguntas y dudas te ayudarán a analizar e identificar los puntos de mejora de los cuales no habías sido consciente hasta la fecha, probablemente porque creías que la gran mayoría de la población tenía un conocimiento y acceso a la tecnología similar al tuyo

Las mediciones de aspectos concretos de la sociedad de la información nos permiten valorar el estado actual en un lugar determinado o en un colectivo de personas y, además, permiten verificar el impacto de políticas concretas de información. Los sistemas de indicadores de este ámbito deben reunir una serie de características que exponemos a continuación:

- El número de indicadores básicos no debe ser excesivo.
- Permite una visión de conjunto de la situación y evolución de la sociedad de la información.
- Incluye indicadores precisos y su definición.
- Tienen una metodología transparente.
- Los indicadores son elaborados a partir de fuentes fiables.
- Son fáciles de interpretar.
- Son susceptibles de comparaciones, incluso a nivel internacional.
- Permiten un desglose de la información por zonas.

Las principales fuentes que debemos tener presente para encontrar estos indicadores son la OCDE y la Unión Europea cuando necesitamos datos internacionales y, en España los diversos observatorios de la sociedad de la información creados por organismos oficiales e informes de larga trayectoria creados por empresas privadas como el realizado por Telefónica.



### PARA SABER MÁS

En la siguiente web puedes acceder a los informes y otras publicaciones periódicas que realiza Telefónica a través de su fundación para evaluar el grado de avance en sociedad de la información.

<https://www.fundaciontelefonica.com/cultura-digital/publicaciones/625/>

Algunos indicadores que debemos tener presentes al analizar la sociedad de la información son:

- Respecto a la industria de las TIC:
  - El mercado de las TIC:
    - Gasto en TIC per cápita.
  - El mercado de las telecomunicaciones:
    - Gasto en telecomunicaciones per cápita.
    - Porcentaje del PIB en telecomunicaciones.
- Respecto a las infraestructuras TIC:
  - % de territorio cubierto con líneas telefónicas.
  - % población abonada a telefonía móvil.
  - % de hogares con acceso a Internet y evolución.
  - % de empresas con acceso a banda ancha y evolución.
- Respecto a los terminales de acceso:
  - % de hogares con ordenador, teléfonos móviles, smartphones, tabletas...
  - % de la población con acceso a ordenadores y evolución.
- Respecto a los servicios:
  - % de la población con acceso a internet de banda ancha y evolución.
  - Número de servidores web y evolución.
- Respecto a los contenidos:
  - % de empresas e instituciones con web.
  - % de webs que permiten comercio electrónico con respecto al total.
  - % medio de presupuesto anual en inversión publicitaria en internet.
  - % de webs según la lengua, sectores...

- Respecto a los usos:
  - Media de consumo por habitante: línea telefónica, móvil, servicios de cable...
  - % de usuarios de banca online con respecto al total de clientes.
  - % de teletrabajadores respecto de la población activa.

Los principales pilares que han favorecido el desarrollo de la sociedad de la información los encontramos a mediados de los años 90 en las políticas de información desarrolladas por la Unión Europea y por el gobierno de Estados Unidos.

En Europa, en diciembre de 1993, nace el libro blanco sobre crecimiento, competitividad y trabajo, más conocido como Libro Blanco de Delors, en honor a Jacques Delors, presidente de la Unión Europea en aquel momento. El objetivo de este libro era definir un espacio europeo que permitiese reactivar el crecimiento económico europeo y la competitividad, así como crear mercados y puestos de trabajo. Se defendía que todo esto solo se podía conseguir a través de la sociedad de la información.



#### PARA SABER MÁS

En la siguiente web de la Unión Europea puedes acceder a informes y publicaciones relativos a la sociedad de la información en cuestiones que tienen que ver con big data, blockchain, ciberseguridad, 5G, etc

<https://ec.europa.eu/digital-single-market/en>

Los objetivos clave del libro blanco eran:

- Difusión de las TIC europeas.
- Modificación del marco reglamentario de las telecomunicaciones.
- Estimular el desarrollo de sistemas telemáticos.
- Efectuar análisis sobre los medios necesarios para que los ciudadanos europeos pudiesen utilizar las nuevas tecnologías.
- Generalizar el uso de las TIC en la enseñanza y la formación.

Al mismo tiempo que en Europa salía a la luz el Libro blanco de Delors, en Estados Unidos, la administración Clinton, diseñaba una política de información federal centrada en los siguientes aspectos:

- La tecnología, porque no hay competitividad sin desarrollo tecnológico.
- La revolución de la información, ya que consideraban que el país que encabezase esta revolución sería el más poderoso.
- La innovación a partir de la tecnología.

Esta política tuvo su punto álgido en 1993, cuando Al Gore anunció el proyecto Nacional Information Infrastructure (NII). Este proyecto pretendía crear una infraestructura nacional de información que tuviese mucha capacidad para el transporte de datos, y que permitiese el desarrollo de aplicaciones telemáticas de altas prestaciones. NII describe la convergencia de las telecomunicaciones, las TIC y la industria del ocio. El objetivo final de este proyecto era fomentar el desarrollo económico e incrementar el liderazgo mundial de los Estados Unidos en este sector. Los principios fundamentales de NII eran los siguientes:

- Fomentar las inversiones privadas en la NII.
- Promover la competitividad.
- Fomentar la innovación y las nuevas aplicaciones tecnológicas.
- Proporcionar un acceso abierto a la NII, tanto para los proveedores como para los consumidores de servicios.
- Generalizar la idea de servicio universal.
- Asegurar la flexibilidad para que el nuevo marco regulador pueda mantener el ritmo de los cambios tecnológicos y del mercado.

En Europa y Estados Unidos se continuó trabajando en el desarrollo de políticas que garantizasen el desarrollo de la sociedad de la información y el acceso a ella por parte de ciudadanos y empresas.

Al Gore, vicepresidente de Estados Unidos en aquel momento, presentó la necesidad desarrollar unas infraestructuras globales de información (Global Information Infrastructure, GII) que requerían la colaboración de todos los países.



#### PARA SABER MÁS

Si quieres conocer la actualidad de las políticas y planes en Estados Unidos a favor de la sociedad de la información puedes visitar esta web:

<https://digital.gov/>

**CITA**

“

“El requisito esencial para un desarrollo sostenible para todos los miembros de la familia humana es la creación de la Red de Redes.

Esas superautopistas o, con más precisión, redes de inteligencia distribuida, nos permitirán compartir información, conectarnos y comunicarnos como una comunidad global.

Desde esta interconexión emanará un progreso económico robusto y sostenible, democracias fuertes, mejores soluciones para los retos ambientales globales y locales, y, finalmente, un mayor sentido de gobierno compartido en nuestro pequeño planeta [...]. El desarrollo de la GII debe ser un esfuerzo cooperativo entre los gobiernos y las personas. No puede dictarse o construirse por una sola nación. Debe ser un esfuerzo democrático.”

Al Gore (1993)

Los principios de este proyecto se presentaron en marzo de 1994, en Buenos Aires, en la ITU World Telecommunication Development Conference. Estos principios eran:

- Alentar la inversión privada.
- Promover la competencia.
- Proporcionar acceso abierto a la red a todos los usuarios y proveedores de información.
- Crear una regulación flexible e inteligente que facilite los cambios rápidos de la tecnología y de los mercados.
- Asegurar el servicio universal.

GII daba un valor económico a la información y hacía más real la ya tópica frase de Bacon “la información es poder” y marcaba el liderazgo de Estados Unidos en el ámbito de la información. Además, se ponía de manifiesto la necesidad de establecer políticas de información nacionales que se encargasen de organizar y regular el marco sobre el que se apoya el proyecto internacional.

En Europa, en junio de 1994, se hizo público el informe de la Unión Europea, Europa y la sociedad global de la información, más conocido como Informe Bangemann (porque fue realizado por el grupo de trabajo encabezado por Martin Bangemann, comisario europeo). Recomendaba a la Unión Europea utilizar mecanismos del mercado para acceder a la era de la información como:

- Fomentar el espíritu empresarial.

- Desarrollar un marco regulador común.
- No invertir más dinero público.

Además, defendía como principales líneas de actuación en las que se debía trabajar:

- Teletrabajo.
- Educación a distancia.
- Redes de universidades y centros de investigación.
- Servicios telemáticos para las PYMES.
- Control del tráfico rodado.
- Control del tráfico aéreo.
- Redes de asistencia sanitaria.
- Redes de concursos públicos.
- Redes entre administraciones públicas.
- Autopistas de información entre las ciudades.

Como consecuencia, la UE decide desarrollar un plan de actuación que permita implantar la sociedad de la información. Para ello, en 1995, la Comisión Europea organizó la Conferencia Internacional sobre la Sociedad de la Información en Bruselas, a la que invitó al vicepresidente de los Estados Unidos, Al Gore, los ministros de industria y telecomunicaciones del G7, una selección de comisarios europeos, así como importantes empresarios del sector privado. Como resultado de esta conferencia, la Comisión aprobó un plan de actuación que centró las actuaciones de la UE en cuatro áreas:

- Desarrollar un marco normativo y jurídico.
- Fomentar la aplicación de las TIC.
- Vigilar y analizar las consecuencias sociológicas, sociales y culturales de la sociedad de la información.
- Promover la sociedad de la información.

A partir de ese momento, la política para la sociedad de la información de la UE queda definida por los siguientes principios:

- Las fuerzas del mercado deben conducir el avance hacia la sociedad de la información. Esto implica una apertura a la competencia en los servicios e infraestructuras de la información.
- Se debe garantizar el servicio universal, la interconexión de redes y la interoperabilidad de los servicios y de las aplicaciones en toda la UE.

- La financiación de la sociedad de la información corresponde al sector privado.
- Se debe proteger y fomentar la diversidad cultural y lingüística.
- El derecho a la intimidad también se debe proteger y fomentar.
- Se debe cooperar con las regiones menos desarrolladas.
- Los operadores económicos deben ser conscientes de las oportunidades que ofrece la sociedad de la información.
- Es necesario sensibilizar al gran público, de manera que deben recibir una formación adecuada.

Desde aquellas fechas, para acelerar el proceso de advenimiento de la sociedad de la información, se han desarrollado por parte de la Unión Europea múltiples políticas.

Esta estrategia se basaba en gran medida en las políticas existentes tanto a escala europea como nacional y tenía por objeto ampliarlas e integrarlas. El plan de acción establecía por tanto objetivos comunes, vinculados a las medidas concretas. La evaluación comparativa de los progresos realizados a escala nacional garantizaría un seguimiento eficaz de tales medidas.

El planteamiento de los jefes de gobierno de la UE en el año 2000 era convertir a la UE en el espacio económico basado en el conocimiento más competitivo y dinámico del mundo.



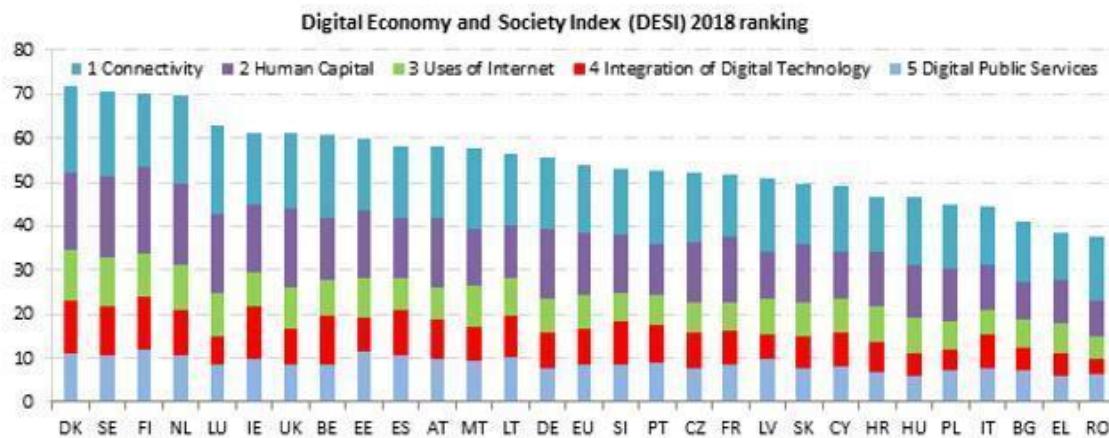
#### COMPRUEBA LO QUE SABES

Desde distintos estamentos se critica la incapacidad de Europa para lanzar al mercado grandes compañías basadas en datos, siendo los referentes del sector estadounidenses (Google, Amazon, Facebook...) o Chinos (Aliexpress, WeChat de Tencent, Baidu, etc.) y Europa solo consigue situar alguna empresa y nunca en las primeras posiciones como Zalando o Spotify. ¿Qué crees que está fallando? Coméntalo en el foro de la unidad.

Realmente, muchos de los planteamientos propuestos no se llegaron a alcanzar o se retrasaron mucho en el tiempo previsto.

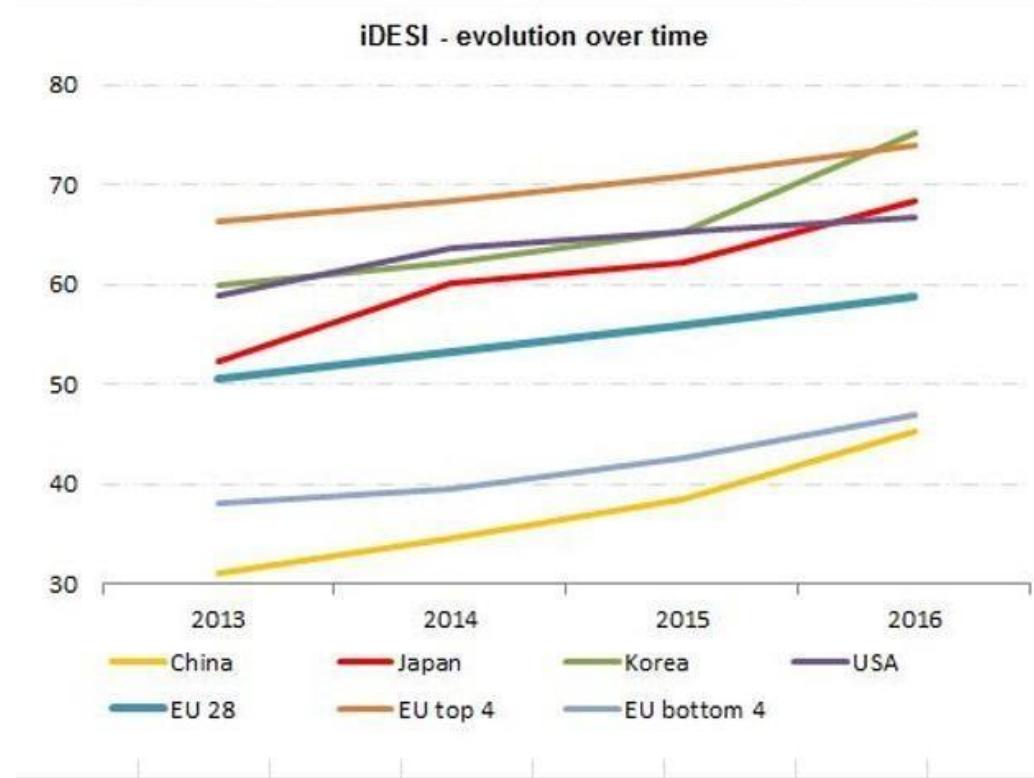
Se buscaba la creación de empleo de forma que Europa se volviese más productiva siendo, además, las condiciones laborales de los trabajadores mejores. Se buscaba también la reducción de la pobreza y una mayor protección del medio ambiente.

Pero ¿en qué posición se encuentra ahora Europa? Estos son algunos indicadores tomados de Eurostat que permiten comparar la evolución de la situación inicial de la UE, los objetivos fijados y la situación de otras potencias económicas.



Clasificación de países de Europa en 2018 según indicadores de sociedad de la información

Fuente: Europa.eu



Comparativa entre Europa y otros grandes países en 2018 acorde al denominado iDESI o Índice de la Economía y la Sociedad Digitales.

Fuente: Europa.eu

Tras las políticas iniciales de promoción de la sociedad de la información, se puso en marcha la estrategia Europa 2020.

**CITA**

“

“Europa 2020 es la estrategia de crecimiento de la UE para la próxima década. En un mundo en cambio, queremos que la UE posea una economía inteligente, sostenible e integradora. Estas tres prioridades, que se refuerzan mutuamente, contribuirán a que la UE y sus Estados miembros generen altos niveles de empleo, productividad y cohesión social.”

Comisión Europea

En aquel momento, la UE fijaba para 2020 los siguientes cinco objetivos:

- Empleo para el 75% de las personas de 20 a 64 años.
- Inversión del 3% del PIB de la UE (sumando pública y privada) en I+D e innovación.
- Respeto a la educación:
  - Tasas de abandono escolar por debajo del 10%.
  - Al menos, un 40% de las personas de 30 a 34 años de edad con estudios de nivel terciario.
- Reducir al menos en 20 millones el número de personas en situación o riesgo de pobreza y exclusión social.
- En cuanto al clima y la energía se pretendía:
  - Reducir los gases de efecto invernadero un 20%, o incluso un 30%, si se diesen las condiciones, por debajo de los niveles de 1990.
  - Un 20% de energías renovables.
  - Aumentar un 20 % la eficiencia energética.

### **1.1.2 La sociedad de la información en España**

*Con toda seguridad, la empresa donde trabajas y estás desarrollando el proyecto comercial para mejorar las ventas se ve afectada por la ley de servicios de la sociedad de la información y de comercio electrónico. Por ello, decides conocer más sobre ella para saber cómo te puede limitar en tu proyecto.*

A partir de las estrategias europeas esbozadas, en España también se desarrollaron una serie de eventos e iniciativas para alcanzar la llamada sociedad de la información.

Muchas premisas iban encaminadas a una mejor regulación del marco legislativo (derecho de propiedad intelectual, privacidad y seguridad) y un perfeccionamiento de todos los recursos, sobre todo humanos. Pero ¿en qué situación se encontraba España en el año 2000 cuando se fijó la Estrategia de Lisboa? Un estudio efectuado a finales del año 2000 nos muestra una sociedad española, con una serie de fortalezas, debilidades, oportunidades y amenazas de cara a afrontar el reto que suponía la sociedad de la información:

- Fortalezas:
  - Facilidades competitivas que ofrece el idioma español (400 millones de hablantes, de las cuales casi 50 millones viven en Estados Unidos).
  - Riqueza cultural.
  - Pertenencia a la Unión Europea.
  - Buen nivel de formación de la población joven.
  - Buen nivel de infraestructura y servicios de telecomunicaciones.
  - Integración en el comercio global.
- Debilidades:
  - Desequilibrios territoriales.
  - Ausencia de cultura telemática en gran parte de la población activa.
  - Bajo consumo per cápita de información.
  - Escaso parque informático instalado.
  - Alto nivel de desempleo juvenil.
  - Debilidad de muchos sectores a la hora de competir en los mercados de las autopistas de la información.

- Oportunidades:
  - La sociedad de la información es un elemento de integración y cohesión económica y social.
  - Posibilidad de utilizar las tecnologías como un elemento de modernización y de aumento de competitividad del tejido empresarial.
- Amenazas:
  - Profundización de desequilibrios de exclusión social y de todo tipo en amplios estratos de la población.
  - Colonización cultural.
  - Deslocalización del empleo.
  - Pérdida de valor añadido propiamente nacional en el nuevo entorno competitivo.

La sociedad de la información conlleva riesgos importantes, que hay que enfrentar aprovechando la posibilidad de utilizar las TIC como un elemento de modernización y competitividad del tejido empresarial, de manera que las empresas españolas puedan competir en los nuevos mercados globales y se utilicen los nuevos soportes y redes para incorporar nuestros contenidos culturales a la sociedad global y ampliar la participación en la cadena de valor de la sociedad de la información.



### EJEMPLO PRÁCTICO

Estás pensando en emprender un negocio online con un grupo de buenos amigos orientado a vender productos de artesanía. Habéis pensado que sería interesante antes de dar el paso, realizar un análisis DAFO de la realidad española y de mercado para tener todo bien pensado.

Tus amigos confían en ti y te encargan esa responsabilidad. ¿Cómo lo harías?

### SOLUCIÓN

El DAFO quedaría de la siguiente manera:

#### Fortalezas

- El mercado de habla española es muy grande, no solo en España, sino que también en países donde el comercio electrónico está muy desarrollado, como Estados Unidos, por lo que es potencialmente muy interesante desarrollar nuestro e-commerce orientado a vender en ese país, que además tiene una capacidad económica para la compra alta.

- También considero que al estar en la Unión Europea y poder vender en todos los estados miembro sin problemas de aduanas, ni aranceles, es una gran fortaleza.
- El nivel de acceso a internet en España es alto y las velocidades de conexión también, con lo cual se podría crear una web muy innovadora para diferenciarlos, con tecnologías de realidad aumentada, para que la gente vea como quedaría en su casa el producto que quieren comprar.

#### Debilidades

- La situación económica por momentos no es la mejor en muchos mercados de habla hispana.

#### Oportunidades

- El sector de la artesanía todavía está muy verde en internet, así que es una clara oportunidad para, que, con nuestros conocimientos de tecnología, provocar una revolución.

#### Amenazas

- Un porcentaje alto de la población ha dejado de valorar la artesanía y prefiere productos foráneos, lo que puede conllevar que nuestros esfuerzos tengan poco crédito, sobre todo al principio y no seamos capaces de sostener el negocio hasta alcanzar la suficiente rentabilidad.

Por otra parte, la tarea de legislar para hacer frente a los retos de la era de la información no es una tarea de fácil solución. Hay que considerar distintas variables que hacen del sector de las nuevas tecnologías uno de los más complejos al incluir un abanico de actividades dedicadas a temas muy diversos.

En la actualidad los sectores económicos que generan conocimiento crecen mucho más rápido que las industrias tradicionales, al igual que los riesgos para los creadores de contenidos que crece exponencialmente cuando los mismos se difunden a través de los medios electrónicos. En este sentido, los mecanismos legales para proteger los contenidos digitales y los mecanismos para asegurar tal protección adquieren especial relevancia.



## NORMATIVA DE INTERÉS

Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia:

<https://www.boe.es/eli/es/rdlg/1996/04/12/1/con>

La copia y distribución a través de internet de productos en formato digital ha constituido desde siempre un tema polémico que aún se debate en distintos tribunales de todo el mundo e incide específicamente en los derechos de autor. El problema afecta a material de todo tipo: software, textos, fotografías, música y otros.



## NORMATIVA DE INTERÉS

Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico.

<https://www.boe.es/eli/es/l/2002/07/11/34/con>

Las organizaciones no gubernamentales e intergubernamentales y los organismos industriales encargados de tratar los distintos aspectos de la falsificación y de la piratería y sus efectos han estimado que el mercado de los productos falsificados e ilegales representa entre el 5 y el 7% del comercio mundial total pero, en esencia, la cuestión radica en que los formatos digitales y la libertad casi sin límites en internet permiten este tipo de hechos y dificultan que las leyes de protección de los derechos de autor puedan hacerse cumplir de forma efectiva.

A eso debemos agregar que, aunque internet es una red que se usa en forma idéntica en todo el mundo, no existe una armonía entre las legislaciones de los distintos países.

En este sentido, la industria audiovisual ha tomado importantes medidas para evitar este tipo de actividades y para que cuando éstas se produzcan se evite que la situación empeore y se impongan sanciones administrativas o penales a sus responsables. De ahí que el Tratado de la OMPI sobre Derecho de Autor (WCT) y el Tratado de la OMPI sobre Interpretación o Ejecución y Fonogramas (WPPT), contengan disposiciones especiales que

prohíben eludir las medidas de protección tecnológica de las obras amparadas por derecho de autor.



#### ENLACE DE INTERÉS

La OMPI es la Organización Mundial de Propiedad Intelectual o en inglés WIPO (World Intellectual Property Organization). Conoce todos sus detalles en:

[www.wipo.int](http://www.wipo.int)

La ley de propiedad intelectual en España contó con el beneplácito de las entidades de gestión de derechos de autor, como la Sociedad General de Autores (SGAE), y las críticas de diversas plataformas que, aunque luchan contra la copia ilegal, se muestran contrarias a esta Ley.

## 1.2 Regulación en materia de protección de datos

*Recuerdas que el director comercial de tu empresa cuando te eligió para el nuevo proyecto te comentó que un proveedor de marketing, probablemente un bróker de datos, le había ofrecido una lista de contactos de personas al parecer supuestamente interesadas en productos de mobiliario como el que tu compañía produce y comercializa.*

Otra de las cuestiones esenciales para que se logre un avance importante en el desarrollo de la sociedad de la información es la confianza de los usuarios en internet. La privacidad se refiere al uso que las empresas hacen de los datos personales que les suministran sus clientes o los usuarios de los servicios.

El aspecto de la privacidad o la protección de datos de carácter personal quedó regulado en España a través de la Ley Orgánica 15/1999, que tiene por objetivo garantizar y proteger, en lo que concierne al tratamiento de los datos personales, las libertades públicas y los derechos fundamentales de las personas físicas, y especialmente de su honor e intimidad personal y familiar.

Esta regulación ha sido el marco de aplicación durante prácticamente dos décadas hasta los recientes e importantes cambios en la materia a nivel de toda Europa debido a la entrada en vigor del GDPR.



### ENLACE DE INTERÉS

Puedes estar al día de todos los cambios normativos en materia de protección de datos desde la siguiente web de la Unión Europea.

[https://ec.europa.eu/info/law/law-topic/data-protection\\_es](https://ec.europa.eu/info/law/law-topic/data-protection_es)

#### **1.2.1 GDPR**

El Reglamento General de Protección de Datos (RGPD), más conocido por su denominación en inglés General Data Protection Regulation (GDPR), es la norma de más alto rango en la Unión Europea en materia de protección de datos personales de las personas físicas y que también regula el libre movimiento de dichos datos dentro de las fronteras de la Unión.

Aunque el reglamento ya entró en vigor el 25 de mayo de 2016, realmente no fue de aplicación plena hasta dos años después, el 25 de mayo de 2018, porque se otorgó ese lapso de tiempo de dos años para que las empresas y todas las organizaciones a quienes le resultase de aplicación pudiesen adaptarse.

Su alcance no afecta solamente a empresas europeas, sino que se aplica a cualquier compañía que quiera desarrollar su negocio en el interior de la UE. Todo dato personal está protegido por este reglamento y las sanciones por su incumplimiento se establecieron muy altas para impedir que en ciertas ocasiones resultase rentable incumplir la ley.

En el momento de entrada en vigor del reglamento, como ya hemos indicado, España contaba con su propia legislación en la materia. Por este motivo, la denominada Ley Orgánica de Protección de Datos de Carácter Personal (LOPD) española del año 1999 se derogó y fue sustituida por la ley de fecha 6 de diciembre de 2018 denominada Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales que es totalmente compatible y coherente con el RGPD.



## NORMATIVA DE INTERÉS

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales:

<https://www.boe.es/eli/es/lo/2018/12/05/3>

Como ya se ha anticipado, uno de los principales cambios que introdujo esta nueva legislación es que el alcance de la ley llega a cualquier empresa del mundo que realice algún tipo de tratamiento sobre datos personales de ciudadanos europeos, las cuales en caso de incumplimiento se verán afectadas por sanciones que pueden llegar a alcanzar el 4% de su facturación a nivel global.

GDPR distingue entre dos tipos de entidades en cuanto a la gestión de datos personales:

- Controlador de datos, que es una empresa que recopila datos de ciudadanos de países miembros de la Unión Europea.
- Procesador de datos, que es una entidad que hace un tratamiento de datos que han sido recopilados por un controlador de datos.

### CITA

“

"Los datos personales son cualquier información relacionada con un individuo, ya sea que se refiera a su vida privada, profesional o pública. Puede ser cualquier cosa desde un nombre, domicilio, foto, dirección de correo electrónico, detalles bancarios, publicaciones en sitios web de redes sociales, información médica o la dirección IP de una computadora".

Comisión Europea

Entre los pocos datos personales que escapan a lo desarrollado en el reglamento están aquellos que tienen que ver con cuestiones de seguridad nacional y algunas otras cuestiones de índole legal o persecución de delitos.

Acorde al reglamento, las reglas en materia de protección de datos son únicas, porque derivan directamente del reglamento, pero cada país tendrá su propia autoridad de supervisión responsable de atender y asesorar a ciudadanos y empresas; investigar usos irregulares de los datos personales; sancionar a los infractores; colaborar con los reguladores de otros países en aquellos casos transfronterizos, etc.

Para facilitar el mercado único europeo, aquellas empresas que desarrollen su negocio en más de un país de la UE solamente responderán ante una autoridad de supervisión principal, que será la del país donde tenga su actividad principal en cuanto a uso y tratamiento de datos.

Desde la autoridad de ese país se supervisará la actividad de la empresa en todos los mercados de la UE bajo la coordinación de un ente denominado Junta Europea de Protección de Datos.

De forma similar a como ocurría por la aplicación de leyes anteriores, las organizaciones que quieran recopilar y utilizar datos personales deben notificarlo a los interesados, incluyendo el tiempo que retendrán sus datos, así como el modo de contactar con el controlador de datos y la información para poder localizar al delegado de protección de datos.



#### ENLACE DE INTERÉS

La autoridad en la materia en España es la Agencia Española de Protección de Datos. En su web puedes recabar información y recomendaciones muy útiles cuando necesites verificar si tu proyecto de business intelligence o big data cumple en materia jurídica:

<https://www.aepd.es/es>

El reglamento define varias bases legales para que una organización pueda hacer tratamiento de datos:

- Cuando es el usuario quien de forma expresa consiente que se realice el tratamiento de sus datos para uno o más propósitos.
- Es necesario tratar unos datos determinados para poder llevar a cabo un contrato donde el usuario es parte o va a ser parte.
- Es necesario tratar los datos para el cumplimiento de una obligación legal por parte del controlador.
- Hay que tratar los datos para poder proteger intereses vitales del titular de esos datos o de otra persona.
- Se requiere el tratamiento de los datos para una tarea de interés público o cuando el controlador de los datos es una autoridad pública.
- Es necesario tratar un conjunto de datos para un interés legítimo del responsable de los datos, excepto que los derechos fundamentales del interesado sean de un rango superior. En caso de que el

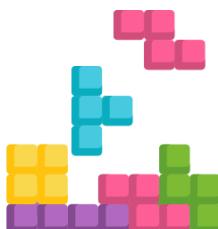
interesado sea un menor de edad entonces se aplicará con mayor sensibilidad, si cabe, esta protección.

Al respecto del consentimiento voluntario, que suele ser el mecanismo más común en cual las empresas se amparan para recopilar datos de todos nosotros, el reglamento recoge que dicho consentimiento tiene que ser explícito, tanto en cuanto a los datos que se pretende recopilar, como en el uso que se les pretende dar. En caso de menores, el consentimiento debe ser otorgado por los padres o tutores.

Los controladores de datos son los responsables de poder demostrar ante terceros que se ha obtenido el consentimiento en los términos expresados. En cuanto al interesado, éste debe saber que puede retirar el consentimiento otorgado en el momento que le resulte conveniente.

Pensemos por ejemplo en las compañías que interaccionan con sus clientes a través de un call center. En base al reglamento, las advertencias tradicionales informando de que la llamada está siendo grabada, ahora no son suficientes como prueba de que el usuario ha dado su consentimiento.

La entrada en vigor del reglamento ha potenciado la figura del delegado de protección de datos. Este rol es obligatorio en las administraciones públicas y en el sector privado cuando el controlador o procesador realiza un tratamiento de datos especialmente intensivo o complejo. La figura del delegado es la de un experto en la materia para garantizar el cumplimiento del reglamento.



### EJEMPLO PRÁCTICO

Estás implantando el sistema de administración y gestión de un partido político, donde entre otras muchas funcionalidades, estás programando el módulo que se encarga de, cada mes, enviar al banco la remesa de recibos para la domiciliación de la cuota mensual de afiliación.

¿Qué aspectos te deberían preocupar más?

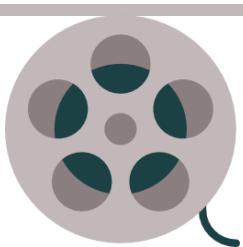
### SOLUCIÓN

Probablemente la decisión técnica más importante que deba tomar es elegir un algoritmo de cifrado robusto, de forma que, en el peor de los escenarios, cuando un hacker robe la base de datos de los afiliados, este no pueda conocer que personas son militantes del partido.

Aparte de mantener la confidencialidad de los afiliados, con ello evitaré tener que notificar de la fuga de información a la agencia de protección de datos y evitaré también sanciones potencialmente muy elevadas.

Aunque no sea responsabilidad mía directa como informático, porque eso depende más del departamento jurídico, sí que es conveniente que informe al responsable correspondiente del partido que por el tipo de datos que se manejan, deben realizar una evaluación de impacto para determinar si están dentro de los rangos asumibles.

Otra figura que crece a raíz del reglamento es el DPO o data privacy officer, que se trata de un rol experto en seguridad de datos, gestión de las tecnologías de la información, ciberseguridad, etc. Que aporta por tanto una visión global sobre cómo proteger los datos y garantizar la confidencialidad de estos, por ejemplo, ante intentos de robo por hackers, etc. Sin duda, el papel del DPO es muy complejo y de máxima responsabilidad.



### VIDEO DE INTERÉS

Si tienes interés en profundizar en el rol del DPO, en el siguiente vídeo encontrarás una interesante ponencia:

<https://www.youtube.com/watch?v=QVrow6fhKZ8>

Cuando se produce en una organización una fuga de información, el reglamento indica que se debe notificar de ello a la agencia reguladora del país y, en caso de riesgo, también habrá que notificar a los afectados para que tomen medidas de prevención oportunas. Siempre que sea posible, la

notificación debe realizarse durante las 72 horas posteriores a haber tenido conocimiento del problema de seguridad.

En caso de que los datos que hayan sido robados o filtrados se hubiesen encontrado cifrados, entonces no sería necesario proceder a las notificaciones indicadas.

Otro recurso que el reglamento introduce para facilitar el almacenamiento y manejo de datos es el de seudonimización.

En cuanto a las sanciones, estas pueden llegar a ser muy elevadas:

- Multa hasta diez millones de euros o la cantidad que resulte de calcular el 2% de la facturación global de la compañía si es más alta.
- Multa hasta veinte millones de euros o el 4% de la facturación global en caso de infringir aquellos aspectos más graves.



#### NORMATIVA DE INTERÉS

REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos):

<https://www.boe.es/doue/2016/119/L00001-00088.pdf>



#### VIDEOTUTORIAL

Aprende a utilizar Facilita RGPD, la herramienta online que la Agencia Española de Protección de Datos pone a disposición de todas las pymes españolas para que comprueben que medidas deben tomar al respecto de los datos que utilizan en caso de que el nivel de riesgo de los mismos sea bajo.

Puedes ver el resultado de esta herramienta en el documento fichero.docx

<https://vimeo.com/user64513894/review/447088073/bd78d69f66>

### **1.2.2 GDPR y big data**

*La empresa que el director comercial de tu compañía ha contratado para comprarles un paquete de datos de potenciales clientes seguramente sean grandes profesionales, pero piensas que no está de más que os preocupéis por verificar si cuentan con el consentimiento de los titulares de esos datos para ser vendidos a vosotros y poder realizar una explotación comercial a partir de ese momento.*

*Es probable que las circunstancias del proyecto que te ha encomendado tu director comercial impliquen la obligación o, cuando menos, la recomendación de realizar una evaluación de impacto de los posibles riesgos asociados.*

*La transparencia será una de las claves del éxito de tu proyecto. Por muy bueno que sea el análisis que realices, sin la debida transparencia lo que puedes generar es confusión o, peor, desconfianza, probablemente el peor enemigo de las ventas.*

Todas las compañías se han visto afectadas por el GDPR, pero aquellas dedicadas a la explotación de datos masivos todavía deben ser más cautelosas para no incumplir la ley, tanto en los procesos de recopilación, como en los de gestión de los datos personales. Al fin y al cabo, como ya hemos mencionado, las sanciones pueden alcanzar el 4% de la facturación anual.

Cuando abordamos un proyecto de big data, debemos tener claro que disponer de un dato no significa que el usuario a quien corresponde dicho dato nos lo haya entregado de forma voluntaria y consciente.

Por ejemplo, ciertos datos pueden tener origen en la observación que realizamos de un usuario apoyándonos en el uso de cookies.



### PARA SABER MÁS

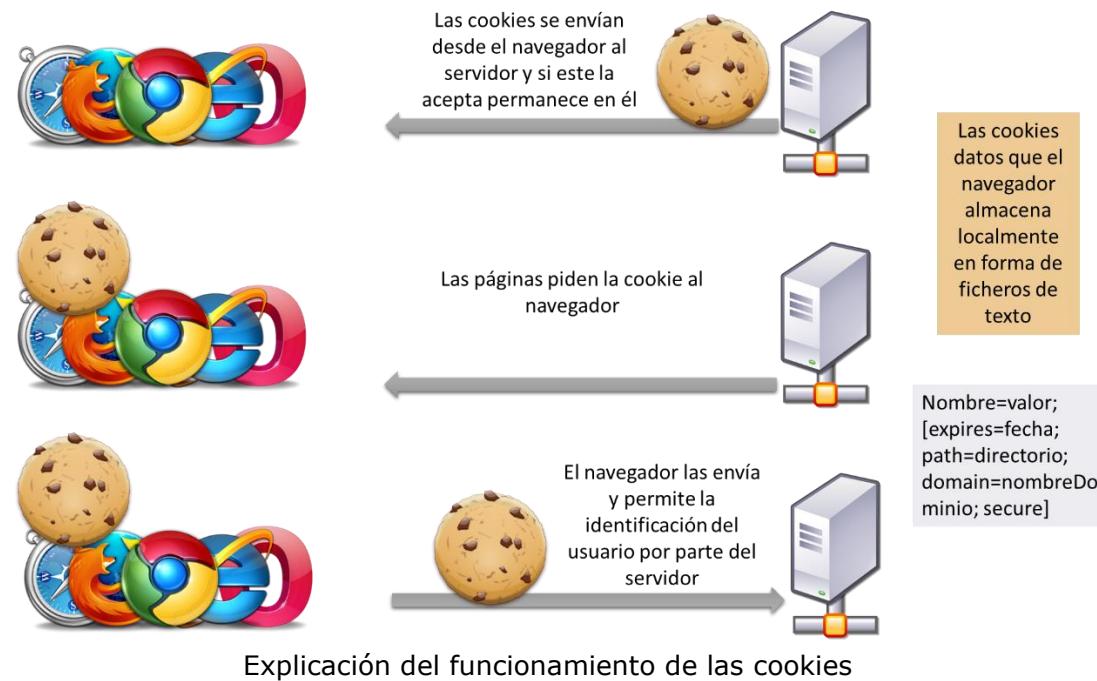
Cuando un usuario se conecta a un servidor de internet, normalmente para visitar una página web a través del browser que haya elegido (Chrome, Firefox, Safari...) el servidor devuelve una cookie, que desde un punto de vista técnico es simplemente un archivo de texto plano como sería un fichero txt.

Este archivo es almacenado en el propio navegador de internet. De esta forma, el propietario del servidor web puede "marcar" al dispositivo del internauta, para reconocer su visita tanto en la sesión abierta en el presente como, potencialmente, en posteriores sesiones que se abran desde el mismo navegador.

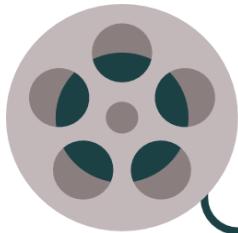
Las cookies, precisamente se desarrollaron para mejorar el concepto de "sesión" que está muy deficientemente gestionado por los protocolos TCP/IP que gobiernan la red.

Aunque es verdad que ese origen y aplicación primigenia, en los últimos años se ha concentrado en un uso de tipo publicitario, lo que ha provocado en cierta medida es una mala imagen de esta técnica, llegando al caso de que muchos usuarios, incorrectamente, consideran las cookies como un tipo de malware.

La legislación de casi todos los países, empezando por los estados miembros de la Unión Europea, son muy sensibles al uso de cookies y por ello obligan a los responsables de las páginas web a notificar al usuario con antelación el uso que van a hacer de las cookies.



Otros datos pueden haber sido obtenidos de fuentes de información accesibles de forma pública.



### VIDEO DE INTERÉS

Los datos abiertos u open data cada vez son más frecuentes. Especialmente administraciones públicas, pero también empresas privadas, comparten sus datos para ser utilizados por otras personas o empresas.

Conoce la experiencia del Ayuntamiento de Málaga en este tipo de iniciativas.

[https://www.youtube.com/watch?v=eT\\_ie4apO-A](https://www.youtube.com/watch?v=eT_ie4apO-A)

Los datos de un usuario, tal vez también puedan haber sido generados a través de un sensor o de cualquier otro elemento componente del internet de las cosas, como pueden ser los datos sobre su ubicación, parámetros de salud, etc.

O incluso, puede ser que dispongamos de datos inferidos, es decir, datos que nosotros mismos hemos sido capaces de crear a partir de otros mediante la aplicación de distintos algoritmos o de mecanismos de machine learning.

En cualquiera de los casos, sea cual sea el origen de estos, ello no nos exime de la obligación de cumplimiento de la legislación en materia de protección de datos.

A la hora de desarrollar un proyecto de big data, para dar cumplimiento a la legislación, debemos optar por la vía del consentimiento o por la de interés legítimo.

También es verdad que algunos de los requisitos del GDPR, como el principio de minimización y el principio de retención de datos, no cuadran del todo bien con proyectos de tratamiento masivo de datos. Para facilitar en ese caso el cumplimiento, es imprescindible definir con claridad el objetivo del análisis desde el momento inicial y garantizar que los datos personales en tratamiento son relevantes para el objetivo que se ha planteado y no son excesivos.



## VIDEOTUTORIAL

La herramienta Gestiona EIPD de la Agencia Española de Protección de Datos te facilitará realizar la evaluación de impacto en la protección de datos personales que en términos de la Agencia es “el análisis de los riesgos que un producto o servicio puede entrañar para la protección de datos de los afectados y, como consecuencia de ese análisis, la gestión de dichos riesgos mediante la adopción de las medidas necesarias para eliminarlos o mitigarlos”.

En este tutorial aprenderás a utilizarla.

<https://vimeo.com/user64513894/review/447090872/488e3491cb>

Un proyecto de big data debe incluir la evaluación de los riesgos e impacto que pueda tener sobre la privacidad de las personas, para que sean contemplados dentro de la gestión general del proyecto. Es lo que se denomina privacidad desde el diseño, es decir, que la aplicación a desarrollar desde el momento inicial se conciba con el cumplimiento normativo en mente. En muchos casos será imprescindible anonimizar los datos.



### SABÍAS QUE...

El concepto privacy by design (privacidad desde el diseño) fue definido por Ann Cavoukian en el 2010.

Este concepto fue asumido como propio por Europa a la hora de desarrollar el RGPD.

Ann Cavoukian es una experta internacional en privacidad y ha sido la responsable de privacidad de la información en el Gobierno de Ontario (Canadá)

De igual modo, es recomendable:

1. Actuar con transparencia cuando nos involucramos en un proyecto big data, de forma que los usuarios sean conocedores del uso de sus datos y conscientes de las implicaciones que ello puede tener. Incluso, más allá del cumplimiento normativo, muchas organizaciones han desarrollado códigos éticos al respecto del uso de datos personales.

2. Para la excelencia en el uso de datos de personas, realizar auditorías internas y externas de los procesos de análisis aplicados a los datos, sean estos procesos humanos o automáticos mediante machine learning u otros algoritmos de inteligencia artificial. El objeto de estas auditorías es que no se acepten algoritmos que discriminen a determinados colectivos en base a un uso sesgado de sus datos.



### ENLACE DE INTERÉS

En este enlace, el profesor del Instituto de Empresa y conocido bloguero Enrique Dans da su opinión sobre un caso de discriminación en Apple por la aplicación de algoritmos de machine learning:

<https://www.enriquedans.com/2019/11/el-machine-learning-y-sus-sesgos.html>

En nuestros proyectos debemos ser cuidadosos a la hora de transformar los datos para evitar perder la conformidad por parte de los usuarios. Por supuesto, los datos de partida deben ser exactos y de calidad, porque de lo contrario el resultado de la algoritmia será inválido.

En aquellos proyectos más complejos es recomendable incorporar al equipo tanto a un experto técnico, como a un experto legal en la materia.

“

### CITA

“Sin duda, la principal forma en que la legislación sobre la protección de datos afectará a la recopilación de datos es que conducirá a una mayor dependencia de los análisis en tiempo real.

El análisis en tiempo real toma los datos que acaban de ser recogidos y los pone en uso y análisis inmediato. Con los datos recogidos no hay necesidad de mantener dichos datos durante mucho tiempo, que es uno de los temas que la GDPR trata de abordar.

Afortunadamente, los investigadores han hecho grandes progresos a la hora de realizar análisis en tiempo real más rápido y más eficaz en comparación con los análisis a posteriori, por lo que hay esperanza de que este cambio en particular se pueda hacer con las mínimas molestias.

Las redes sociales, una vía que muchas empresas utilizan con el fin de crear lealtad en los clientes y aumentar el compromiso, también se verán

afectadas. Esto no es sorprendente, considerando cuánta información personal termina residiendo en las cuentas de los medios sociales.

Si observamos la reciente noticia de que Facebook ha liberado nuevas herramientas de privacidad, se hace evidente que las reglas están cambiando.

Las empresas que quieran hacer negocios con clientes de la UE tendrán que ser más cuidadosas con lo que piden y ser más francas con respecto al tiempo que retienen esos datos y lo que harán con ellos.

Además, todos esos grandes datos que se están recogiendo no sólo tendrán que ser almacenados de forma segura, sino que tendrán que ser entregados a los clientes que quieran eliminarlos y cambiarlos a otro proveedor. Las empresas de todos los tamaños tendrán que aceptar la idea de que los clientes tendrán un mayor control sobre sus propios datos personales.

Así que el reloj está corriendo, y se espera que se cumpla. Las grandes organizaciones tienen los recursos necesarios para ayudar a reducir el impacto en cuanto al cumplimiento de la ley de protección de datos.

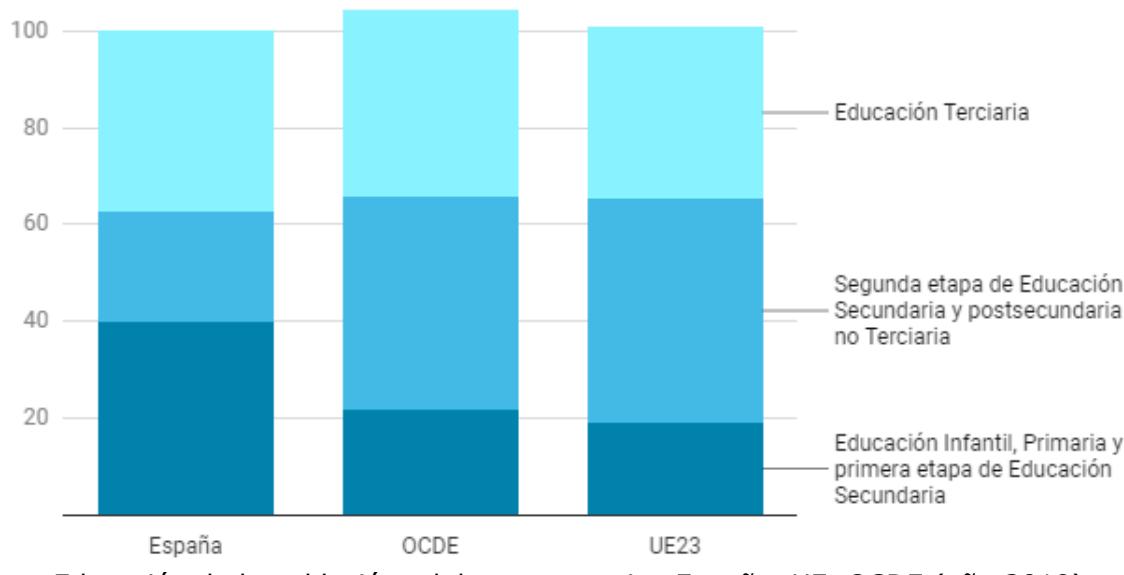
¿Pero qué pasa con las pequeñas y medianas empresas? ¿Cómo prosperarán con estas nuevas regulaciones?

Traducido de un original de John Terra. Simplilearn.com

### **1.3 Los recursos humanos en la sociedad de la información en España**

En cuanto al desarrollo de los recursos humanos, hay que distinguir varios aspectos imprescindibles que deberían tomarse en cuenta para poder aprovechar al máximo la sociedad de la información: nivel educativo en España y porcentaje de especialistas en TIC.

A pesar de los esfuerzos por elevar el nivel educativo de la población activa española, ésta todavía tiene un nivel de formación inferior a la media de la OCDE.



Educación de la población adulta comparativa España, UE, OCDE (año 2019)

Fuente: Instituto nacional de evaluación educativa

Además, se caracteriza por una alarmante polarización, que se plasma en altos porcentajes de población con niveles educativos extremos: unos bajos y otros con titulación universitaria, mientras que sólo existe un porcentaje muy reducido de población con formación de segundo ciclo de secundaria o educación post-secundaria no universitaria.

El problema no es la abundancia de titulados universitarios (donde, aunque se sobrepasa la media de la OCDE, no se alcanzan los niveles de Estados Unidos, Dinamarca, Austria o Hungría), sino el escaso atractivo de la formación no universitaria para los jóvenes españoles.

Las titulaciones técnicas son escasas. España es de los países con menor proporción de estudiantes matriculados y titulados en ciencia y tecnología. Este es el aspecto más negativo de la enseñanza superior española para la competitividad empresarial, su capacidad de innovación y su velocidad de adaptación a las nuevas tecnologías.

No vamos a entrar en detalles acerca de todas las políticas o planes aprobados en España, que de alguna manera están relacionadas con la información o el desarrollo tecnológico, porque esto supondría el análisis de la evolución de la situación española en este ámbito desde los años 80, momento en el que surgen las primeras iniciativas relacionadas con las políticas de información, marcadas por la influencia de políticas externas.

Sin embargo, consideramos importante tener presente cuáles han sido los hitos que, de una manera u otra, han contribuido a configurar la situación española actual en lo que se refiere a la sociedad de la información.

<b>1982</b>	Creación de la Subdirección General de Documentación e Información Científica
<b>1983-1986</b>	Plan IDOC
<b>1986</b>	Ley de la Ciencia. Plan Nacional de Investigación Científica y Desarrollo Tecnológico
<b>1997</b>	Ley General de Telecomunicaciones
<b>1999</b>	Comisión Interministerial de la Sociedad de la Información
<b>2000</b>	Creación del Ministerio de Ciencia y Tecnología
<b>2001-2003</b>	Plan Info XXI: La sociedad de la información para todos
<b>2002-2003</b>	Comisión Soto
<b>2003</b>	Plan de choque para el impulso de la administración electrónica en España
<b>2004-2005</b>	España.es
<b>2004-2007</b>	Plan Conecta
<b>2006-2008</b>	Plan Moderniza
<b>2006-2010</b>	Plan Avanza
<b>2010-2015</b>	Plan Avanza 2
<b>2015-2020</b>	Estrategia nacional de ciberseguridad
<b>2020</b>	Plan Nacional 5G

El primer plan que se aprobó en España, en el cual se tiene en cuenta el papel fundamental de la información en la sociedad, es el Plan IDOC (Directrices para un Plan Nacional de Actuación 1983-1986 en materia de Documentación e Información Científica y Técnica). En él se recomendaban, entre otros aspectos, las siguientes cuestiones:

- Desarrollar una infraestructura de difusión de la información nacional e internacional.
- Desarrollar los recursos humanos necesarios para los servicios de información.
- Sensibilizar a la sociedad del valor de la información.
- La formulación de políticas de información por parte de las administraciones públicas.

Aunque este plan partía de buenos principios para facilitar la inclusión de los españoles en la sociedad de la información, la realidad es que su repercusión fue muy baja, al no haberse diseñado un plan con dotación presupuestaria.

La Ley de Ciencia de 1986 (Ley 13/86 de Fomento y Coordinación de la Investigación Científica y Técnica) tenía los siguientes objetivos:

- Progresar en el desarrollo y la innovación tecnológica.
- Desarrollo de los servicios públicos.
- Conservación del patrimonio artístico e histórico.

- Progreso y difusión de la cultura.
- Adecuación a las TIC.

Establecía, además, un Plan Nacional de Investigación Científica y Desarrollo Tecnológico, coordinado por la CICYT (Comisión Interministerial de Ciencia y Tecnología).

Con la llegada de los años 90 se produce una globalización de la economía a nivel mundial, aceleración del cambio tecnológico, necesidad de nuevos modelos educativos y formativos, disminución del protagonismo del sector público, transformación del mercado del trabajo... y España, que ya estaba integrada en la UE, comienza a participar de las políticas europeas sobre las TIC que surgen como consecuencia de la conferencia del G7 de 1995. Como consecuencia de la adopción de las directrices europeas, en 1997, se aprueba la Ley 12/1997, General de Telecomunicaciones que liberaliza el sector de las comunicaciones. En 1999 se crea la Comisión Interministerial de la Sociedad de la Información, que tiene como objetivo incluir a España en la sociedad de la información a partir de la coordinación con otras instituciones y las comunidades autónomas y, en el año 2000, se crea el Ministerio de Ciencia y Tecnología, que engloba la política científica y la política de información a través de una Secretaría de Estado para las Telecomunicaciones y para la Sociedad de la Información, que incluye una Dirección General para el Desarrollo de la Sociedad de la Información.

En el año 2001 se aprueba el primer plan con medidas específicas para el desarrollo de la sociedad de la información: Info XXI. Info XXI fue, ante todo, una acción política muy polémica. A grandes rasgos, se trataba de un plan de acción, con gran multitud de actividades, que tenía como objetivo introducir y potenciar la sociedad de la información en España eliminando la brecha digital, consiguiendo que cualquier individuo radicado en España pudiese acceder a la sociedad de la información en los múltiples roles que pudiese adoptar en su vida cotidiana: empresario, trabajador, estudiante, ciudadano, etc.

La polémica asociada a este plan se debe, fundamentalmente, a dos motivos. En primer lugar, Info XXI fue presentado y posteriormente modificado y, en segundo lugar, la polémica se debió a la dotación presupuestaria del mismo ya que, aunque en términos absolutos no estaba mal dotado (aproximadamente 825.000 millones de pesetas), en realidad, solo 280.000 eran de inversión real, ya que las partidas restantes estaban destinados a gastos corrientes de los diversos ministerios y administraciones beneficiarias del plan o, bien, en Info XXI se habían integrado presupuestos relacionados con las TIC que ya estaban previamente asignados. Por ejemplo, las compras de equipos ofimáticos para funcionarios por renovación del parque, el mantenimiento de servicios

de administración electrónica que ya se encontraban previamente en marcha, etc.

Info XXI, en lo bueno y en lo malo, no puede negar que es hijo del proyecto eEurope y, por ello, no tuvo muy en cuenta las particularidades y realidades de la sociedad española, con dificultades tan marcadas frente a los vecinos europeos que eran un lastre excesivo para el plan: escasa cultura de innovación en el tejido industrial y de servicios español, mayor porcentaje de población rural sin acceso mínimo a la red, etc.

España.es, programa surgido como resultado del informe de la Comisión Soto, tuvo una financiación aproximada de 1.029 millones de euros (de los que la Administración General del Estado (AGE) aportó el 63%, las Comunidades Autónomas el 26% y el sector privado el 11%).

Este programa tuvo una duración de dos años (2004 – 2005), y apoyó la cultura de la innovación tecnológica a través del Plan Nacional de I+D+I. El programa se centró en tres grandes líneas de actuación:

- Reforzar la oferta de contenidos y servicios que favorezcan la demanda de Internet.
- Mejorar la accesibilidad en sentido amplio, ofreciendo puntos de acceso público y haciendo un esfuerzo en formación y comunicación de las ventajas de la sociedad de la información.
- Conectar a la pequeña y mediana empresa, aumentando su relación de negocio a través de internet, con el fin de que pueda acceder a servicios de la sociedad de la información.

Esas tres líneas se desglosaron en seis grandes actuaciones:

- Administracion.es: Impulsar la administración electrónica.
- Educacion.es: Internet en la escuela II. Del “aula de informática” a la “informática en el aula”.
- Pyme.es: Incorporación de las pymes a la sociedad de la información.
- Navega.es: Acceso de los ciudadanos a la sociedad de la información.
- Contenidos.es: Contenidos digitales de calidad.
- Comunicacion.es: Comunicar las ventajas de la sociedad de la información.

Dentro del plan España.es encontramos el plan de choque para el impulso de la administración electrónica en España. Este plan abordaba cuatro frentes de actuación:

- Facilitar el acceso público a los usuarios.
- Impulsar el desarrollo de servicios para los usuarios.

- Facilitar el intercambio de información entre las AA.PP.
- Apoyar la reorganización interna de procesos de las AA.PP.

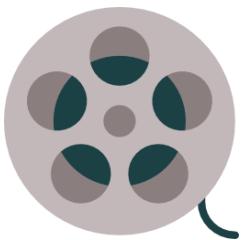
La inclusión de las AA.PP. en la sociedad de la información dio un paso más con el desarrollo del Plan Conecta (2005-2007). Los objetivos de este plan eran:

- Mejora de la calidad.
- Agilidad y rendimiento de los servicios a los ciudadanos.
- Eficiencia en el uso de los recursos públicos.
- Reducción de costes.
- Satisfacción de los usuarios.
- Integración interdepartamental.
- Simplificación administrativa.

Y las principales líneas de actuación de este plan fueron:

- CERTIFICA. Que trataba de reducir las visitas que el ciudadano debe realizar a la administración.
- eDNI. Se intentó fomentar la implantación de la firma electrónica en la tramitación de procedimientos administrativos y facilitar la incorporación del DNI electrónico.
- Ciudadano.es. Pretendía ofrecer medios de comunicación y servicios avanzados interactivos a través de múltiples canales, facilitando la participación ciudadana en el diseño de las políticas públicas.
- Simplifica. Era necesario configurar una gestión pública racional y eficiente eliminando barreras y cargas burocráticas que lo entorpezcan.

Uno de los ejes clave de la profundización de la sociedad de la información en España fue el Plan Avanza. El Ministerio de Industria, Turismo y Comercio de España elaboró un Plan para el desarrollo de la Sociedad de la Información y de Convergencia con Europa (Plan Avanza) y entre comunidades autónomas y ciudades autónomas, orientado a conseguir la adecuada utilización de las TIC para contribuir al éxito de un modelo de crecimiento económico basado en el incremento de la competitividad y la productividad, la promoción de la igualdad social y regional y la mejora del bienestar y la calidad de vida de los ciudadanos.



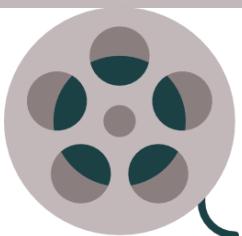
### VIDEO DE INTERÉS

Presentación del informe más reciente sobre la sociedad de la información en España por parte de la Fundación Telefónica.

[https://www.youtube.com/watch?v=EzmIntbvW6Y&feature=emb\\_logo](https://www.youtube.com/watch?v=EzmIntbvW6Y&feature=emb_logo)

Pero, ¿en qué situación nos encontramos en este momento? El último informe sobre la sociedad de la información en España de la Fundación Telefónica, publicado en 2019 y con datos de 2018 -principalmente- nos muestra como los aspectos más destacables de la sociedad de la información española los siguientes:

- El 5G ya está aquí. Aunque la nueva generación de telecomunicaciones no acaba de ser todavía una realidad general a nivel de mercado, sí que ya está claro que impactará en todos los ámbitos de la vida y que el plan nacional en 5G que ha puesto en marcha el gobierno va a potenciar su asimilación.



### VIDEO DE INTERÉS

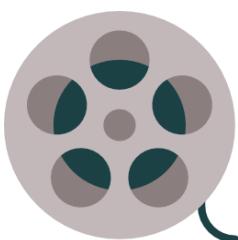
Vodafone nos explica su visión de cómo el 5G transformará el mundo también en lo que tiene que ver con el big data:

<https://www.youtube.com/watch?v=7f-ddrH6Ilw>

- Los asistentes de voz como Amazon Echo o Google Home están cambiando como las personas interactuamos con equipos y dispositivos.
- El GDPR ha provocado cambios significativos en cómo se gestionan los datos y en el valor que se le da a la privacidad por parte de las empresas. También está cambiando la redefinición de los conceptos alrededor de los derechos de autor.
- La inteligencia artificial está extendiendo y generalizando su uso tanto en los hogares como en las empresas y los bots, chatbots, voice bots son cada vez más comunes. Ahora es el momento de que la sociedad se convenza de los aspectos positivos de la inteligencia

artificial y pierda miedo a la creencia de que ha venido para destruir puestos de trabajo.

- La ética en el uso de los datos, sobre todo a raíz de los escándalos como Cambridge Analytica y el impacto que tuvo en distintas contiendas electorales alrededor del mundo ha sensibilizado a toda la sociedad sobre los efectos que puede tener la analítica y el big data sin criterios morales.

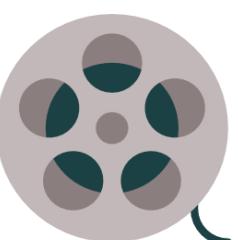


#### **VIDEO DE INTERÉS**

El trabajador de Cambridge Analytica que reveló la trama, afirma que sin el uso que hicieron del big data político Reino Unido todavía seguiría dentro de la Unión Europea:

<https://www.youtube.com/watch?v=xwSP0PYL8Vo>

- El sector financiero reclama un nuevo marco regulatorio que les ofrezca más flexibilidad para poder experimentar nuevos productos y servicios, como aquellos basados en blockchain, de forma que puedan competir en igualdad de condiciones contra los nuevos servicios fintech.



#### **VIDEO DE INTERÉS**

Ponencia patrocinada por Orange sobre la relación entre blockchain y big data

<https://www.youtube.com/watch?v=mq2-Sk9NmS0>

- El edge computing empieza a asentarse como un enfoque de futuro, especialmente por su vinculación al internet de las cosas y a la nube. Este tipo de computación lo que busca es un equilibrio entre la computación tradicional y la nube, de forma que ciertas capacidades de procesamiento estén junto al origen de los datos y no en la nube.
- Los nuevos riesgos y problemáticas relacionadas con la seguridad de la información, como por ejemplo la creación de redes de ordenadores zombis por parte de los hackers para realizar minería de criptomonedas.
- La consolidación total del cambio de hábitos de consumo de ocio, especialmente en lo que tiene que ver con la migración de la

televisión a las plataformas tipo Netflix o Amazon Video, que sostienen en el big data y el análisis avanzado su capacidad para conocer al consumidor y poder crear contenidos audiovisuales en base a la demanda.

Penteo, una reconocida consultora y analista de mercados tecnológicos, en su informe más reciente, con menos de un año de antigüedad, destaca que:

- El ERP todavía es el software más importante en cuanto a compras por los departamentos de informática.
- La migración a la nube se ha ralentizado, tal vez por la incertidumbre ante la explosión del edge computing.
- El gobierno y la gestión operativa de las tecnologías de la información toma importancia, recibe más presupuesto en las empresas y se aspira a disponer de una visión más estratégica.
- En cuanto al uso de los datos e información, el mayor foco por parte de empresas y todo tipo de organizaciones está en reducir los silos de información.
- Los proyectos de transformación digital se apoyan tanto en tecnologías tradicionales que se han renovado, como los CRM, junto con tecnologías emergentes.
- Las operaciones en materia de ciberseguridad son cada vez más críticas, sobre todo en lo que tiene que ver con la prevención del impacto de los ataques de ransomware.

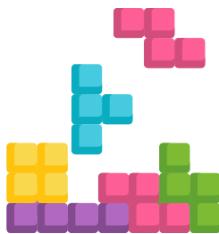
Finalmente, para tener una visión de las tendencias tecnológicas para 2020 y años sucesivos, nada mejor que conocer el informe de la más conocida e influyente consultora de mercados tecnológicos, que es Gartner.

Cada año, la consultora Gartner anuncia cuáles son las diez tecnologías estratégicas que marcarán el año y que deben ser tenidas en cuenta por todas las empresas. Para el año 2020, estas son las tecnologías que destaca:

- Hiperautomatización, mediante la combinación del machine learning y las herramientas de automatización, de forma que el aprovisionamiento de recursos informáticos y la prestación de servicios sea más ágil. A la automatización de procesos robóticos (RPA) se unirán otras herramientas que incrementen su potencia.
- Experiencia multisensorial y multimodal, que quiere decir que los usuarios experimentarán cambios en cómo perciben el mundo digital, así como en el modo que interactúan con él. La voz es un ejemplo claro del cambio en las interfaces, pero también se generalizará el uso de realidad virtual y realidad aumentada, que conforman la realidad mixta.

- El conocimiento tecnológico se democratiza, de forma que profesionales sin conocimientos TIC avanzados podrán explotar tecnologías complejas, como el aprendizaje automático, el análisis avanzado o el desarrollo de aplicaciones. Específicamente, Gartner destaca que esta democratización será más clara en:
  - Explotación de datos y análisis, ya que, con la mejora de las herramientas dirigidas a científicos de datos, estas serán usables por otros perfiles.
  - El desarrollo de aplicaciones que hagan uso de la inteligencia artificial estará al alcance de muchos, al no ser necesario escribir código para crear estos aplicativos.
  - Acceso al conocimiento, que permitirá a profesionales no IT, explotar su conocimiento de negocio de forma más sencilla.
- El “superhumano”, es decir, el uso de tecnologías para mejorar las capacidades cognitivas y físicas de las personas empieza a vislumbrarse como una realidad plausible. En cuanto al incremento de capacidades físicas, éstas se basan en el concepto de cyborg, es decir, la integración de elementos externos en el cuerpo humano. En cuanto a la mejora cognitiva, ésta viene del mayor acceso a fuentes de información, del incremento de la potencia de las aplicaciones que explotan esa información y de la mejora en las interfaces.
- Más transparencia y trazabilidad en pos de una ética digital y del respeto a la privacidad, porque los consumidores en general somos más conscientes del valor que tiene proteger nuestra información y datos personales. Del mismo modo, las empresas son cada vez más sensibles al impacto negativo que puede tener en su marca, prestigio y cifra de negocios el uso inadecuado de los datos. Además, con el previsible incremento en la aplicación de la inteligencia artificial a los datos, las cuestiones éticas tienen más importancia que nunca.
- Edge computing, entendido como una arquitectura de sistemas mediante la cual, sin renunciar a los beneficios de la nube, se pretende procesar los datos en la ubicación más próxima a donde son generados. De esta forma, con la reducción de los flujos de datos se reducen costes, se rebaja el impacto de la huella medioambiental que generan los grandes centros de datos remotos y se reduce la latencia. Gartner considera que la explosión definitiva del IoT o internet de las cosas conllevará la consolidación del modelo Edge, porque el coste de desplazar a la nube tal cantidad de datos no es eficiente, y más si tenemos en cuenta que muchos de los objetos que estarán conectados al internet de las cosas son críticos, como el vehículo autónomo, y un incremento de la latencia al enviar sus datos a la nube o un mínimo retraso en el procesamiento de un conjunto de datos, puede tener consecuencias desastrosas.

- La nube distribuida consiste en que la nube pública se lleva a diversas ubicaciones, de modo que los servicios que ofrece pasen a encontrarse distribuidos. Al tiempo, el proveedor de la nube, como pudiese ser Amazon o Microsoft, se responsabiliza de coordinar las operaciones, actualizaciones y cualquier otra tarea requerida para un correcto funcionamiento.
- Objetos autónomos que, apoyándose en la inteligencia artificial, no requieren de un humano para la realización de determinadas tareas de forma inteligente. Es decir, van más allá de una automatización básica, lo que hace que los percibamos como inteligentes. El vehículo autónomo será el objeto de este tipo que tendrá más impacto en la sociedad, siempre que la legislación lo permita y los humanos lo aceptemos y nos acostumbremos a ello. Incluso, la interacción entre esos objetos autónomos permitirá alcanzar unos resultados de mucho mayor impacto.
- Uso del blockchain en aplicaciones diferentes de las criptomonedas, lo que transformará la forma de funcionar de muchos sectores económicos e industriales. Blockchain aportará sus atributos principales, como son la confianza y la transparencia, lo que cambiará muchas relaciones empresariales, reducirá costes, eliminará intermediarios, etc. Aportará seguridad y confianza a muchos sectores donde la trazabilidad no ha llegado o en los cuales se desconfía de quienes se supone que son los garantes de la seguridad. Por ejemplo, facilitará conocer el origen de cualquier alimento o evitar la comercialización engañosa de productos falsificados.
- Cambios en cómo se gestionan las identidades, también apoyados en blockchain y en los contratos inteligentes.
- Avances en la ciberseguridad, gracias a la inteligencia artificial y al aprendizaje automático. Pero al mismo tiempo, también se incrementarán las brechas de seguridad, por ejemplo, las que tendrán origen en la generalización del internet de las cosas y en la necesidad, por tanto, de hacer seguros los miles de millones de dispositivos de todo tipo que estarán conectados a la red.



### EJEMPLO PRÁCTICO

Un buen amigo que es propietario de un despacho de abogados te pide consejo sobre cuáles crees tú que son los aspectos tecnológicos que más le van a afectar en el futuro para poder estar preparado para ello.

### SOLUCIÓN

Tú le comentas que son múltiples los factores que, a la vista de las opiniones de los principales analistas de mercado, van a impactar sobre su sector.

Probablemente el blockchain sea uno de ellos, porque va a cambiar como se realizan los contratos de todo tipo.

También le informas que los riesgos informáticos son cada vez más altos y, en el caso de su despacho, debería estar muy prevenido frente a los ransomware, que pueden llegar a secuestrar la información de todos sus clientes, lo que le generaría una situación de la que sería prácticamente imposible recuperarse.

Igualmente, le trasladas tu opinión de que los RPA crees que van a sustituir muchos procesos realizados actualmente por personas, así que probablemente en su despacho puedan reducir muchos costes administrativos y de "papeleo" si sustituyen personal administrativo por RPAs.

## 1.4 Los sistemas de BI

La inteligencia de negocio o BI -por sus siglas en inglés- business intelligence, es un modelo para la gestión estratégica de las organizaciones, el cual está enfocado a explotar los hechos relevantes contenidos en la información histórica con el objetivo de generar un alto valor añadido que, en principio, puede parecer intangible pero que, a lo largo de esta unidad, veremos cómo puede convertirse en una herramienta fundamental para crear valor y mejorar los ingresos.

La inteligencia de negocio surgió ante la necesidad creciente de las organizaciones que requerían constantemente de métodos más eficientes y efectivos para realizar análisis de información a lo largo de su cadena de valor.

El BI cubrió esta necesidad ya desde inicios de la década de los sesenta cuando por primera vez se utilizó el término de inteligencia de negocio por uno de los grandes actores del mercado de la tecnología que fue IBM.

Uno de sus investigadores, Peter Luhn, introduce el término en su artículo “A Business Intelligence System”, donde establece que business intelligence es “la habilidad de aprender de las múltiples relaciones de los hechos en un entorno determinado, que permitan orientar las acciones y decisiones hacia una meta definida”.

De esta década datan también los primeros usos de discos rígidos o discos duros, que eran principalmente usados con fines de investigación en computadoras que ocupaban grandes cuartos y tardaban días enteros en el procesamiento de la información.

Diez años después, alrededor de 1970, nacen los sistemas gestores de bases de datos (SGDB), usando modelos relacionales, que permitían la interacción con los datos objeto de estudio, facilitando la creación de consultas y la definición de estructuras de datos, así como reducían los ciclos de explotación, entre otros asuntos. Edgar Codd fue quien realizó la descripción inicial de este sistema relacional de datos.

Como resultado de esto, sumado al crecimiento acelerado de la globalización y al incremento exponencial de las tecnologías emergentes de la mano de las primeras versiones de internet, en los inicios de los ochenta se crean los primeros ordenadores para un uso comercial y cómodo de SQL.

Todo ello, junto con el aumento de las capacidades de procesamiento y almacenamiento de datos, sienta las bases de un concepto más avanzado que veremos a lo largo de esta unidad que es el data warehouse o almacén de datos.

Los data warehouse fueron la primera piedra de una marcada tendencia de crecimiento o evolución hacia los sistemas de soporte de decisiones o DSS (decissions support systems).

Pero no es hasta principios de la década de los noventa cuando Howard Dresden, un analista de la firma más grandes de investigación en tecnología, Gartner Group, propone una definición estructurada para un concepto que era algo abstracto hasta el momento y lo define como: “conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos”.

Después de ese momento y con el lanzamiento de Excel 1.0 al mercado, la definición sufre distintos cambios y transformaciones, de la misma forma que las tecnologías emergentes se volvían más accesibles al empresario promedio.

El cambio de siglo trajo novedades en cuanto a la explotación de los datos, sobre todo la facilidad -inimaginable tiempo atrás- para acceder a múltiples plataformas y a sistemas altamente eficientes para el análisis de datos, ya no solo a un nivel estructurado, sino también a datos no estructurados. Nos estamos refiriendo a la arquitectura de datos cliente-servidor, que permitía disponer las bases de datos distribuidas, mejorando no solo la calidad de la información capturada, sino la complejidad, alcance y escalabilidad potencial.

#### ***1.4.1 Inteligencia de negocios en la era de la transformación digital***

En este nuevo milenio, la hiper conectividad de internet se hizo más que evidente, poniendo al alcance de la mano, con bajas inversiones, herramientas muy potentes para el análisis multivariado de datos, basándose en modelos de estadística descriptiva, modelos predictivos y variadas técnicas de optimización. Por primera vez en la historia, la explotación de los datos y su análisis se volvieron fichas clave en el juego empresarial. Las reglas y los mercados son cada vez más dinámicos y saber usar bien estas piezas es clave para la configuración de un estrategia competitiva y diferenciadora.

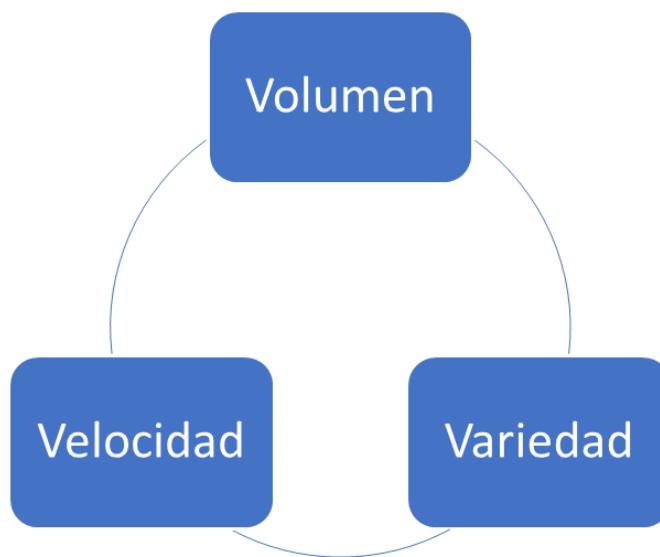
Dave Rogers, de Boston Consulting Group, enumera lo que para él son los cinco dominios para liderar la transformación digital:

- Datos: Es el valor aprovechable, basado en hechos, para realizar predicciones.
- Innovación: Es determinante aprender y experimentar de forma acelerada.
- Competencia: La era digital facilita la entrada de nuevos competidores, de aquí que sea clave crear estrategias competitivas.
- Valor: Es el beneficio final que va a percibir el cliente.
- Clientes: Los clientes en un entorno empresarial son los que cambian y exigen los cambios, por tanto, anticiparse a sus necesidades puede ser clave en el crecimiento corporativo.

### 1.4.2 Cómo entender el business intelligence en tiempos de big data

El concepto de business intelligence se ha visto redefinido en los últimos tiempos por una de las grandes tendencias de la transformación digital que es el big data.

En ocasiones se identifican los datos que explota el business intelligence con los grandes datos o macrodatos del big data, pero la realidad es que no son lo mismo.



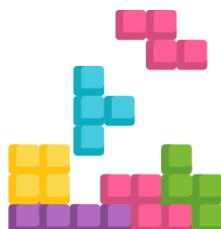
Características del big data

Fuente: Elaboración propia

No todo el business intelligence es big data. Para ser considerado como big data, los datos a explotar deben presentar varias de las siguientes características:

- Volumen: Derivado de la hiperconectividad del mundo actual, se generan cada minuto volúmenes muy altos de datos que deben ser almacenados de forma óptima.
- Velocidad: El estar inmersos en un mundo globalizado y el surgimiento de tecnologías como el internet de las cosas, provoca que aumente la velocidad con la que se generan los datos en distintas fuentes.
- Variedad: La variedad en los datos no radica únicamente en las fuentes de origen de estos, sino que tiene que ver con lo variada que puede ser su naturaleza. Estamos hablando desde un simple número de cuenta bancaria, hasta datos que pueden obtenerse del patrón de una persona al caminar o correr.

En toda estrategia de business intelligence es determinante controlar los datos y contemplar esas tres perspectivas, pues como comprobamos a lo largo de estas unidades el impacto en los negocios es alto, lo que lleva a que en los próximos años todas las empresas deberán desarrollar sus capacidades, no solo de almacenar de forma óptima los datos, sino también de traducirlos e interpretarlos en la búsqueda de más beneficios.



### EJEMPLO PRÁCTICO

Estáis charlando entre tres compañeros y uno de ellos comenta que está trabajando en el departamento de mantenimiento de una central eólica y que todo lo que hace es big data. Por lo contrario, el otro amigo le dice que no está de acuerdo y que eso no es big data.

El primero insiste en que los aerogeneradores están constantemente monitorizados. Que cada segundo se registra múltiples parámetros como la velocidad de giro de las hélices, la velocidad del viento, la temperatura del aceite y una larga lista más. Que, del mismo modo, la red de distribución genera datos en tiempo real que deben ser capturados 20 veces por segundo para evitar cualquier sobrecarga de tensión, etc. Etc.

Se enzarzan en una discusión por ver quien lleva la razón y como no llegan a un acuerdo, te piden a ti que des tu opinión para dirimir la disputa.

### SOLUCIÓN

Tu les explicas a ambos que las características del big data son:

- El volumen y parece claro que eso se cumple, dado se generan gran cantidad de datos nuevos de forma continua lo que unido a la gran cantidad de dispositivos que tiene una compañía eléctrica obliga a almacenar terabytes de datos cada día.
- La velocidad también es una característica que está presente en base a lo descrito, dado que los datos se toman con alta frecuencia, incluso algunos de ellos, como ha explicado el compañero, varias veces por segundo.
- La variedad, por lo contrario, parece no estar presente en los datos gestionados, ya que la naturaleza de ellos parece muy similar, al tratarse todos de mediciones físicas como temperatura, velocidad de rotación, presión, kilovatios, etc. Pero, por lo contrario, no se están tomando datos de imágenes, ni de sonidos, ni de redes sociales, ni de documentos...

Así que, siendo justo, tienes que darle la razón al amigo que opinaba que no se trata de un caso de big data.

### **1.4.3 ¿Cuándo se requiere la inteligencia de negocios BI?**

Como se ha indicado, la inteligencia de negocios es una tendencia que cada vez toma más fuerza de la mano de la tecnología, dando pasos de gigante que nos llevan a una evolución de las capacidades organizacionales nunca vistas.

Pero, a pesar de ellos, a muchos responsables empresariales les surgen duda como, por ejemplo: ¿cómo se si necesito business intelligence? ¿Cuándo es el momento oportuno para usarla? ¿Qué síntomas o características deben considerarse para optar por su aplicación?

La naturaleza de las situaciones que pueden llevar a una organización, en un sector determinado, a implantar business intelligence puede ser tan variable y diferente que en cualquier caso se requiere un conocimiento profundo del negocio y un análisis concienzudo de los procesos, usuarios, empleados, maquinas o sistemas usados.

Resaltaremos a continuación una serie de elementos comunes a todas las empresas que pueden ser considerados, en cualquier caso. En términos orgánicos, todas las empresas poseen eso que se denomina "cultura organizacional", que es en pocas palabras el conjunto de valores, percepciones, actitudes, hábitos, formas de interacción existentes e, incluso, modelos de toma de decisiones, que la enfocan a lograr el cumplimiento de la estrategia. Por este motivo, algunas empresas son más proclives a una toma de decisiones intuitiva, mientras que otras aplican modelos más formales.

Otro aspecto a analizar en cualquier caso es el posible uso inapropiado de ciertas herramientas conforme al crecimiento de la empresa. Muchas organizaciones llevan años invirtiendo en tecnologías relacionadas con los datos lo que las puede haber llegado a disponer de gran cantidad de herramientas y tecnologías heredadas que en la actualidad lastran más que aportan.

También es común, en muchas organizaciones, que se hayan creado y operado repositorios de información a lo largo de toda la cadena de valor lo que ha derivado en un caos de información, pues la creación de múltiples versiones y versiones de documentos provoca diferencias de información entre áreas y procesos, que aumentará considerablemente el coste de unificarlas.

Por otro lado, como se ha visto anteriormente, el acelerado crecimiento de la tecnología e internet, impulsan a las empresas a adaptarse a los cambios por lo que surge la necesidad de realizar cruces rápidos de información

entre áreas, lo que a su vez genera un aumento exponencial de datos, lo que lleva a un volumen de información imposible de analizar con métodos tradicionales.

De igual modo, las fuentes de información se vuelven más complejas y diversas, como, por ejemplo, el uso de cámaras de seguridad, dispositivos biométricos, redes sociales, etc. Por lo tanto, realizar una extracción, transformación y carga de datos se vuelve complejo.

#### **1.4.4 *Como definir la estrategia de business intelligence***

La definición de una estrategia apropiada de business intelligence no es algo que surja espontáneamente de la noche a la mañana y el proceso de desplegar todo el modelo en la organización tampoco es sencillo de realizar. Es imperativo que la dirección esté altamente comprometida con la creación del modelo, dado que ello requiere se realice una integración de tecnologías, usuarios, personas, procesos y las formas en que se hacen las cosas porque de otro modo será imposible alcanzar un modelo sólido y riguroso.

La implementación del modelo de business intelligence es un trabajo a largo plazo, que requiere esfuerzos, retos y compromisos a todos los niveles de la empresa. A continuación, resaltaremos algunos elementos a tener en cuenta para iniciar la implantación del modelo de inteligencia de negocio:

- Identificar los procesos de negocio que requieren o necesitan aplicar técnicas de análisis de forma continua.
- Definir un BSC o Balanced Scorecard, que es básicamente un esquema de indicadores de desempeño y gestión del rendimiento de la empresa en sus procesos clave, basado en el cumplimiento de los objetivos estratégicos.
- Establecer e identificar claramente los datos e información relevante y suficiente que se va a tratar y/o a involucrar en la minería de datos.
- Definir y elegir a los líderes de proceso o encargados que incidan de forma determinante en la organización.
- El sponsor debe estar abierto a planear, considerar y realizar inversiones tácticas en tecnologías de la información que permitan adquirir tecnologías de referencia y herramientas competitivas.
- Todo cambio lleva consigo la aparición de brechas, miedos e inestabilidades entre los involucrados. Por lo tanto, se debe realizar una labor aristotélica para educar y formar a la organización en pro de la mejora y al tiempo que se gestiona adecuadamente este cambio.

- Elegir metódicamente las necesidades que se pretenden cubrir con la inteligencia de negocio, pues como suele ocurrir en muchos proyectos no se priorizan las necesidades basándose en el valor que van a generar, sino en los usuarios que más protagonismo tienen. Por lo anterior, elegir bien es clave en el camino del business intelligence para generar valor.
- Alinear al departamento de sistemas y tecnologías de la información con la estrategia del negocio y los demás departamentos, para que se trabaje de forma integrada, orientados al logro y a tener una solución a la medida.

Claramente, hasta este punto, vemos que la implementación de un modelo de business intelligence establece un proceso de transformación -no solo a nivel cultural- sino a nivel tecnológico, estratégico e institucional para enfocarse en datos y que de estos emanen las decisiones. Como veremos a lo largo de estas unidades, esto es lo que realmente podemos denominar como una empresa data driven.

#### ***1.4.5 ¿Cómo saber en qué nivel de la inteligencia de negocio se encuentra mi organización?***

Teniendo ya un primer contexto de lo que significa el business intelligence, así como las razones que definen su estrategia, las ventajas o beneficios aportados y el esquema general de cómo abordar la inteligencia de negocios, es importante también saber en qué estado se encuentra la organización.

Aquí presentaremos un modelo predefinido, pero que debe ser estudiado con detenimiento y sobre el cual daremos más detalle a lo largo de estas unidades. Este tipo de conocimientos puede resultar especialmente útil para aquellos que trabajen en oficinas técnicas, oficinas de proyectos, departamentos de tecnología o departamentos de BI o, incluso, para quienes son emprendedores de su propio negocio.

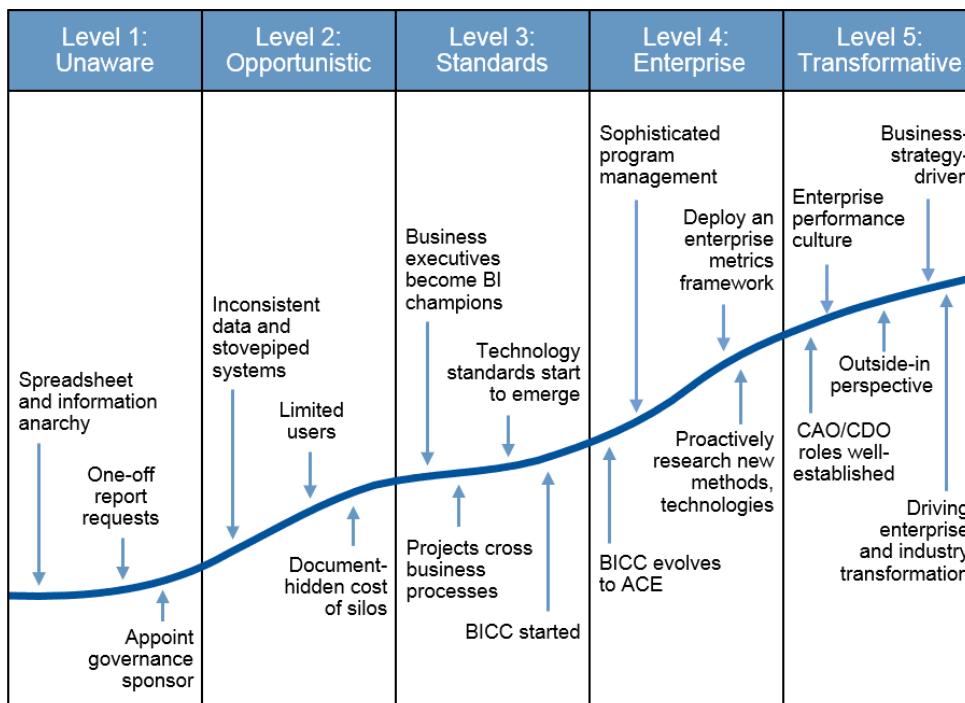
Este modelo del cual estamos hablando se conoce como BIMM (Business Intelligence Maturity Model) y fue desarrollado por Gartner Research y a continuación explicamos cómo funciona, basándose en una escala de cinco etapas o niveles que indican cómo de madura está nuestra estrategia de business intelligence.

- Business intelligence no existe: La información se haya desagregada y distribuida a lo largo de la organización de forma anárquica, en repositorios personales, lo que genera diferencias de información,

lentitud e insuficiencia de datos, que lleva a la toma de decisiones intuitiva.

- Oportunidad de business intelligence: No existe un proceso formalizado para el tratamiento de los datos, ni su acceso. Solo unos cuantos usuarios pueden acceder a información de calidad y consistente. El uso excesivo de Excel y otras herramientas ofimáticas hace todavía más presente las islas de información, al mismo tiempo que lleva a los usuarios más avanzados a la suposición de la existencia de herramientas más efectivas las cuales no son facilitadas por la organización. No hay un modelo de inteligencia de negocios, se crean silos de información y cuellos de botella del procesamiento de la información, lo que genera costes ocultos y difíciles de detectar.
- Estándares de proceso de business intelligence: Los estándares iniciales emergen por la tecnología subyacente a este modelo. La organización establece un equipo encargado de centralizar y controlar la información, que se puede definir como un BICC (Business Intelligence Competency Center). Asimismo, en muchas ocasiones, la extracción de la información se realiza mediante un proceso no modelado y directamente de los sistemas transaccionales, sin realizar limpieza o curado de los datos.
- Modelo empresarial avanzado: Se evidencia el impacto negativo de los repositorios virtuales deslocalizados, lo que crea la necesidad de emplear sistemas sofisticados de almacenamiento y se ve al data warehouse como una solución idónea. Asimismo, se ha desarrollado un sistema estructurado de métricas e indicadores empresariales.
- Transformación a business intelligence: Se ha implantado una cultura empresarial enfocada al logro y al desempeño basado en la información y los datos. Se implementa una estrategia de negocio impulsada por datos (data driven) en los procesos más significativos de la organización, generando impactos positivos que ya se hacen evidentes, lo que conlleva a una transformación organizacional. De ese modo recurrente, surgen nuevas necesidades de hacer procesos más eficientes y modelos más efectivos para analizar y procesar información. La organización entiende la diferencia entre OLTP y DSS.

Aunque existen muchos modelos para el análisis de madurez del modelo de inteligencia de negocios, la herramienta de Gartner es una de las más recomendadas y utilizadas. A continuación, se presenta una imagen ilustrativa del modelo.



Niveles del BIMM

Fuente: Gartner.com

Como se ha presentado, el modelo de madurez de inteligencia de negocios BIMM ofrece únicamente una aproximación a cómo evaluar y determinar el estado actual del modelo de BI. Pero, una vez abordado el proceso de madurez, el modelo debe ser mejorado constantemente, basado en la comparación con los competidores y cómo igualar sus competencias y mejorarlas.



#### PARA SABER MÁS

Accede a los artículos de Melody Chien si quieres profundizar en estas cuestiones.

<https://www.gartner.com/analyst/62722/Melody-Chien>

Pero entonces, ¿cómo usar el modelo de madurez en beneficio de la empresa sin que se sobrevalore ni pierda su valor?

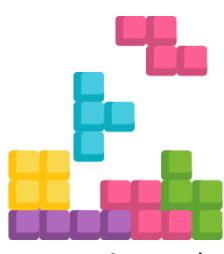
Una analista de Gartner Institute, Melody Chien, propone tres simples pasos para abordar más eficazmente el modelo.

1. Establecer una estrategia de corto plazo: La analista propone que se deben establecer objetivos de corto plazo, en el término de un año máximo, entorno a la estrategia de análisis de datos.
  - Se deben establecer hitos y tiempos de realización que sean razonables y que puedan alcanzarse en ese lapso de tiempo. Esta estrategia debe enfocarse en pequeños incrementos, que produzcan beneficios tangibles y que puedan demostrar a la empresa el valor que el sistema de inteligencia de negocio puede ofrecer en el manejo de los datos.
  - Los incrementos rápidos se enmarcan en lo que Gartner llama el modo 2 para datos y análisis, que es un enfoque ágil y rápido de cómo funcionan las pequeñas empresas emergentes.
  - Garter considera que es la forma de cómo debería funcionar la estrategia de business intelligence, lo que permite un desarrollo gradual de competencias y ganancias rápidas que generan un mayor impacto a nivel comercial.
  - Ello no quiere decir que debamos permitir que la estrategia de corto plazo sea nuestro único horizonte, dado que esos incrementos o ganancias rápidas deben aprovecharse para ir a por los objetivos de largo plazo de nuestra estrategia de análisis de datos.
  - Finalmente, hay que vincular estrechamente no solo la inversión de compra de un software de BI, sino también el motivo de compra de éste y estar convencido de su valor. De lo contrario, solo serán inversiones desperdiciadas.
2. Constituir un equipo de inteligencia de negocios: No se trata propiamente de un equipo en sí mismo, o de que se requiera una estructuración jerárquica de un departamento de BI. Más bien se propone la estructuración de un equipo multidisciplinar e interdepartamental que involucre a participantes de todas las áreas de la empresa.  
Este equipo se organizará alrededor de objetivos y no en base a roles específicos. Ellos serán los responsables de configurar la estrategia de BI y ponerla en marcha. Tienen como propósito fundamental asegurar el programa de business intelligence para que este satisfaga las necesidades de todos los departamentos de la organización, de forma que el resto de los usuarios puedan acceder y manejar la información y toma de decisiones basadas en datos.
3. Establecer un modelo de gobierno de datos escalable: Establecer un gobierno de datos comienza con la determinación e identificación efectiva de los procesos involucrados, sus datos e interacciones y como, como ya se ha indicado anteriormente, tomando en consideración las tres V: volumen, velocidad y variedad.

Igualmente se debe establecer un método seguro para garantizar que los datos están limpios, son precisos, usables y seguros.

No tener un gobierno de datos definido a priori provocará que la búsqueda de soluciones a posteriori resulte más compleja y costosa. No solo el sistema de BI debe estar alineado, también todos los sistemas transaccionales de la compañía, lo que permite que los usuarios accedan a la información que necesiten y no a un montón de datos innecesarios.

Los analistas de Gartner, asimismo, recomiendan que el gobierno de datos no debe ser algo impuesto o restrictivo. Por lo contrario, debe ser un consenso de las influencias que se ejerza en la organización. El desarrollar este gobierno de forma colaborativa hará que los involucrados vean esto de forma cooperativa y no se opongan a ello.



### EJEMPLO PRÁCTICO

Te han pedido asesoramiento en una pequeña cadena de supermercados de tu familia, que quieren comenzar a desarrollar un proyecto interno de business intelligence para mejorar el negocio.

Ellos no saben qué pasos dar para y confían en tu consejo ¿qué les propondrías?

### SOLUCIÓN

En base al modelo definido por Gartner, podríamos definir tres fases claras:

- En primer lugar, definir una serie de pasos a corto plazo, que podrían ser, conocer qué productos se venden mejor según el perfil del comprador, como evolución las ventas según el día de mes o de la semana, qué incremento de ventas se consigue con cada tipo de promoción desarrollada, etc.
- En segundo lugar, buscaría entre los departamentos de la empresa, probablemente marketing y finanzas serían los más interesantes, a un grupo de empleados para crear un equipo de trabajo y avanzar en la materia.
- Finalmente, tendríamos que definir un modelo de gobierno de los datos, con el objetivo principal de mejorar la calidad y fiabilidad de los mismos, para que vaya mejorando y los análisis de BI realizados sean cada vez más precisos y poderosos.

#### 1.4.6 Retos que afrontar

Como estamos viendo en esta introducción, el business intelligence y el crecimiento de los datos supone para las organizaciones la necesidad de desarrollar la habilidad de administrar y gestionar la información sin riesgos. Esto requiere establecer lazos de confianza inquebrantable con el cliente interno y los involucrados claves, creando una cultura colaborativa alrededor de los datos y los hechos que estos cuentan a través de interpretaciones estadísticas de los mismos.

Como se desprende de muchas ideas ya expresadas en esta unidad, las sociedades desarrolladas están experimentando en los últimos años una revolución -entendida ésta como un proceso de cambios vertiginosos- debida a múltiples aspectos, tanto sociales como económicos, políticos y tecnológicos. Algunas de las principales causas de esta revolución, que podemos enumerar, son las siguientes:

- Globalización de los mercados, mediante políticas coordinadas para la reducción de trabas a los movimientos de personas, mercancías y capitales.
- Competitividad creciente y reducción de márgenes comerciales.
- Tecnologías emergentes que fomentan la productividad, facilitan la comunicación entre las personas y las entidades, y permiten el acceso a niveles de información cada vez mayores y desconocidos hasta la fecha, lo que ha supuesto una democratización en el proceso de obtención de información por parte del público en general.
- Consumidores y clientes más y mejor informados, con niveles de exigencia cada vez mayores.
- Ciclos de vida de los productos cada vez más cortos, que exigen a las compañías una constante renovación de su oferta de productos.

Estas y otras cuestiones configuran el nuevo escenario donde las organizaciones deben desarrollar sus actividades y al que deben adaptar sus estrategias empresariales si quieren crecer o, simplemente, sobrevivir. Algunas fórmulas para combatir con éxito las crecientes exigencias y aprovechar las oportunidades que el mercado ofrece son las siguientes:

- Reducción del time-to-market, es decir, del tiempo transcurrido desde que se detecta una nueva necesidad en el mercado hasta que la organización pone a disposición de la red comercial el producto o servicio que la satisface. Las primeras empresas en mover ficha dispondrán de una ventaja competitiva sobre las rezagadas.
- Orientación a la demanda o, dicho de otro modo, adecuación de la estrategia de producción y comercialización a los productos y

servicios más solicitados por el mercado o que, aun siendo incipientes, presenten una clara curva de evolución ascendente.

- Reducción de costes, mejorando el proceso completo (investigación, producción, aprovisionamiento, distribución, comercialización, etc.), y asumiendo que éste presenta ineficiencias.
- Utilización de nuevos canales de distribución y comercialización, algunos de ellos como consecuencia de la aparición de nuevas tecnologías (directo, mayorista, minorista, agentes, oficinas, telefónico, internet...).
- Campañas de marketing más eficaces y precisas, dirigidas a un público más selectivo y con mayor predisposición al consumo del producto publicitado, con el objetivo de aumentar la captación de nuevos clientes.
- Mejora de la calidad del servicio de atención a clientes, incorporación de nuevos servicios, suministro de más y mejor información, con el objetivo de satisfacer sus exigencias cada vez mayores y aumentar su fidelidad.

Las citadas fórmulas, entre otras, permiten a las organizaciones ser más competitivas y adaptarse al nuevo escenario. Para desarrollarlas es necesario, además de otras consideraciones que exceden al ámbito de las tecnologías de la información, proporcionar a los equipos directivos y gestores de la compañía una plataforma de suministro de información de gestión, con características de calidad, rapidez, comodidad y sencillez, que les permita conocer el comportamiento de su negocio en términos absolutos y en relación con el sector, además de facilitarles la toma de decisiones más ajustadas a la realidad de la compañía.

## 2. ARQUITECTURA DE SISTEMAS DE BI

*En el proyecto que estás trabajando, el director comercial será el principal cliente de la aplicación, debes desarrollar el objeto para que pueda decidir qué ofertas comerciales va a remitir a cada lead tomando en cuenta las características de los mismos.*

Business intelligence (BI), o inteligencia empresarial, es el conjunto de instrumentos utilizados en una organización con el objetivo de facilitar a sus gestores las tareas de planificación, control de gestión y toma de decisiones. Estos instrumentos dan lugar a sistemas de información construidos específicamente para apoyar el proceso de toma de decisiones y que genéricamente se han denominado sistemas de soporte a la decisión o DSS.

Las soluciones de business intelligence para el apoyo a la toma de decisiones son muy variadas, lo mismo que sus aplicaciones. Como suele decirse, un sistema de soporte a la decisión no se compra, sino que se construye. Esta afirmación no es del todo cierta, pues existen algunas excepciones que evitan un desarrollo propio:

- La mayoría de ERP's (Enterprise Resource Planning) o softwares integrados para la gestión y planificación de distintas áreas de la empresa disponen de sus propios módulos de análisis de información.
- Soluciones estándar o aplicaciones especializadas, normalmente orientadas a un solo departamento, que responden a modelos de negocio perfectamente estandarizados y de carácter general; dentro de esta categoría se encuadrarían las aplicaciones de planificación y presupuestación, control de riesgos, gestión de balances, etc.

Mencionamos estas excepciones porque es conveniente que se conozca la existencia de estas alternativas, a pesar de que nos centraremos a continuación en desarrollar y analizar el conjunto de elementos que intervienen en la configuración de una solución a medida de las necesidades de cada organización.

El concepto de business intelligence se manifiesta en sistemas de información dirigidos a los equipos gestores de las compañías, analistas del negocio y alta dirección, diseñados específicamente para apoyar el proceso de toma de decisiones.

El objetivo, por tanto, será el establecimiento de una plataforma de suministro de información que facilite el proceso de toma de decisiones a través del conocimiento del comportamiento del negocio. La mencionada plataforma debe respetar los siguientes requisitos:

- La calidad de la información proporcionada debe estar garantizada, siendo éste el elemento fundamental para permitir una toma de decisiones correcta. La información insuficientemente actualizada, incorrecta, incompleta o inútil invalidará el sistema.
- La cantidad de información suministrada debe ajustarse a la estrictamente necesaria para cada tipología de usuario y necesidad concreta. El exceso de información distrae la atención del usuario de los aspectos más importantes, dificultando el análisis y la toma de decisiones.
- La información debe desplegarse en términos de negocio, es decir, debe ser comprensible por los usuarios que la van a utilizar
- La información debe fluir a todos los niveles decisarios de la organización, lo que permitirá descentralizar el proceso de toma de decisiones a los diferentes niveles de responsabilidad.
- La información debe desplegarse en plazos útiles para su utilización, y de forma inmediata cuando sea el usuario quien realice la solicitud.
- Los mecanismos para el acceso a esta información por parte de los usuarios deben ser sencillos de utilizar.
- En determinadas aplicaciones, la solución debe disponer de mecanismos activos, capaces de detectar e identificar por sí mismos aquellas contingencias programadas o previstas, e informar puntualmente a los interesados.

En definitiva, el objetivo de la solución es la creación de un sistema de información que proporcione a sus usuarios (directivos, gestores y analistas de negocio) unos mecanismos sencillos y cómodos de utilizar para la obtención de información de gestión, ajustada a cada usuario y necesidad concreta, con criterios de calidad, precisión, concisión y claridad. Esta plataforma se utilizará como medio para la obtención del conocimiento del comportamiento del negocio, lo que permitirá un proceso de toma de decisiones eficaz y más ajustado a la realidad de la organización.

## 2.1 Antecedentes

*En tu proyecto tienes claro que una fuente muy importante de datos es la que te proporciona un proveedor o bróker externo que está en disposición de venderte un conjunto de perfiles de potenciales clientes.*

De todo lo antedicho, queda suficientemente claro que no siempre será suficiente con aplicar una solución paquetizada, sino que sus contenidos son altamente dependientes de cada organización y de sus intereses. No

obstante, la solución sí presenta ciertos estándares y elementos comunes que vamos a analizar.

Aunque los términos business intelligence, sistemas de soporte a la decisión, data warehouse, datamart, etc. son relativamente modernos, su filosofía es bien conocida desde bastante tiempo atrás. Las organizaciones siempre han manifestado la necesidad de disponer de información de gestión para acomodar sus estrategias empresariales a la realidad del entorno.

Tradicionalmente, las organizaciones han centrado sus esfuerzos en la mecanización de los procesos que les permitían hacer funcionar el negocio. Los sistemas de información tradicionales, como los aplicables a recursos humanos, producción, distribución, facturación, almacenes, contabilidad, y un largo etcétera, así como los actuales ERP's, son claros ejemplos de aplicaciones que abordan el núcleo del negocio de las compañías. Estas aplicaciones se conocen popularmente como sistemas operacionales, en contraposición a los sistemas decisionales (DSS).

A medida que estas necesidades van siendo satisfechas, crece la demanda de sistemas de información cuyo énfasis no es hacer funcionar el negocio, sino analizar el comportamiento de este, en aras a perfeccionarlo y alcanzar mayores cotas de desarrollo empresarial. La materia prima de estos sistemas la constituyen los datos manejados por los sistemas operacionales, si bien estos no constituyen las únicas fuentes.

Hasta hace relativamente poco tiempo, las organizaciones se apoyaban exclusivamente en los sistemas operacionales para el suministro de información de gestión, con mejoras puntuales en ciertos aspectos del proceso de obtención de información. Algunas soluciones primitivas que han sido implementadas son las siguientes:

- Poner a disposición de los usuarios finales una serie de herramientas diseñadas específicamente para la utilización por éstos, con el propósito de facilitarles la obtención de información desde los sistemas operacionales, reduciendo la dependencia de las TIC y agilizando el proceso. Efectivamente, las peticiones a los departamentos TIC de nuevos informes de gestión tardaban mucho tiempo en ser atendidas, y para cuando el informe estaba generado, había perdido su utilidad. Esta técnica es popularmente conocida como "data warehouse virtual", donde los modelos de información permanecen invariables.
- Abordar un proceso de reingeniería para centralizar en un único almacén de datos toda la información disponible relativa a un determinado elemento de gestión (también conocido como entidad), típicamente el cliente, evitando su dispersión. Esta técnica se conoce

como "reingeniería centrada en entidad", y persigue aumentar la comprensión del sistema por parte del usuario.

- Incorporar nuevos recursos hardware y software de mayores prestaciones para mejorar los tiempos de respuesta en la obtención de información. Se trata de aumentar el rendimiento de la plataforma para obtener resultados más rápidos, mejorando la percepción del usuario.

Todas estas soluciones son "remedios de urgencia" que no pueden considerarse como auténticos sistemas que faciliten la toma de decisiones. Se apoyan exclusivamente en los sistemas operacionales, los cuales fueron construidos para soportar la mecanización de los procesos que hacen funcionar el negocio y están dirigidos, normalmente, a usuarios de oficina, pero son ineficaces para soportar el proceso de toma de decisiones por parte de los directivos, equipos gestores y analistas de negocio.

Proporcionan datos brutos, pero no información, pues los modelos de datos no fueron concebidos con este propósito. Los datos se encuentran dispersos por las diferentes aplicaciones, y su recuperación y consolidación resulta una tarea muy ardua. Sin embargo, constituyen la fuente de información más valiosa de la compañía, si son debidamente estructurados y sometidos a un proceso de transformación y depuración que los convierta en información útil.

El siguiente cuadro recoge las principales debilidades que presentan los actuales sistemas operacionales, y en contraposición, qué debe caracterizar a los sistemas de información para que puedan ser utilizados como un instrumento eficaz en el proceso de toma de decisiones:

Sistemas operacionales	Sistemas decisionales
Desde el punto de vista de la información	
Modelos de datos complejos para los equipos gestores, diseñados para la mecanización de procesos, ajenos a la toma de decisiones, recogiendo multitud de datos irrelevantes.	Modelos de datos comprensibles por los equipos gestores, diseñados específicamente para el proceso de análisis del negocio y para apoyar la toma de decisiones.
Multitud de sistemas, desarrollados a lo largo del tiempo, sin integración, dispersos por diferentes plataformas, percibidos como "islas de información" que dificultan su explotación.	Repositorio común con toda la información necesaria, integrada y consistente; procesos unificados de obtención de información (cada dato se obtiene por un único proceso).
Datos inconsistentes, incompletos, erróneos, redundantes, no utilizados...	Excelencia en la calidad de la información para permitir la toma de decisiones correctas.
Pobreza de datos históricos e inexistencia de datos de entorno (mercado, competencia, información sociodemográfica, etc.).	Información histórica rica para establecer evoluciones y tendencias, así como de entorno para establecer correlaciones y permitir la investigación.
Desde el punto de vista de la explotación	
Explotación determinista de la información, con un diseño y construcción previos, altamente dependiente del departamento de informática y dirigida fundamentalmente a usuarios de oficina.	Explotación determinista y también libre, con mecanismos que permiten la explotación de los sistemas por los equipos gestores de forma cómoda, sencilla, rápida y libre, con la única limitación del propio modelo de datos.
Sistemas de información y estructuras de datos ineficientes para el análisis, al estar orientados al registro de sucesos, es decir, al procesamiento transaccional on-line (OLTP).	Sistemas de información y estructuras optimizadas para las consultas y el procesamiento analítico on-line (OLAP).

#### Comparativa sistemas operacionales y decisionales

Fuente: Elaboración propia

El concepto moderno de business intelligence incorpora todos los elementos que pueden ser utilizados por la organización para disponer de un DSS

acorde a sus necesidades. Normalmente se utilizan todos o algunos de los siguientes aspectos:

- **Implantación del modelo de negocio**, que incluye todos los objetos básicos de la actividad empresarial (personas, productos, puntos de venta, contratos, operaciones,...), todas las relaciones entre los mismos, otros objetos construidos por composición de los objetos principales (estructuras), el diccionario de datos donde residirá la codificación, descripción, origen y métodos de cálculo de todos los datos almacenados en el repositorio central, así como los metadatos de negocio, que establecen la correspondencia entre los conceptos de negocio manejados por los usuarios en la explotación y los correspondientes datos físicos, aumentando la comprensión del sistema por los usuarios finales.
- **Infraestructura de datos especializada**. Los modelos entidad-relación han demostrado su eficacia en entornos operacionales, orientados a la transacción. No obstante, estos modelos se muestran menos eficaces en entornos decisionales, donde prima la consulta sobre grandes volúmenes de información. Para determinadas tipologías de explotación, es más conveniente utilizar modelos multidimensionales, pues la información se almacena en estructuras apropiadas para ser explotadas por los diferentes ejes relevantes para el análisis.
- **Hardware dedicado al entorno decisional**. Es una solución más flexible y racional, que permite seleccionar la mejor plataforma y los mejores productos para las necesidades planteadas, además de proporcionar independencia entre entornos, lo que asegura que no van a interferirse mutuamente. El entorno decisional va a requerir una explotación antagónica en muchos aspectos al operacional, pudiendo comprometer el rendimiento de los actuales sistemas operacionales que se encuentran dimensionados para una cierta carga de trabajo y a una tipología de explotación. En la siguiente tabla se recogen las principales divergencias de ambos entornos, desde el punto de vista de la explotación:

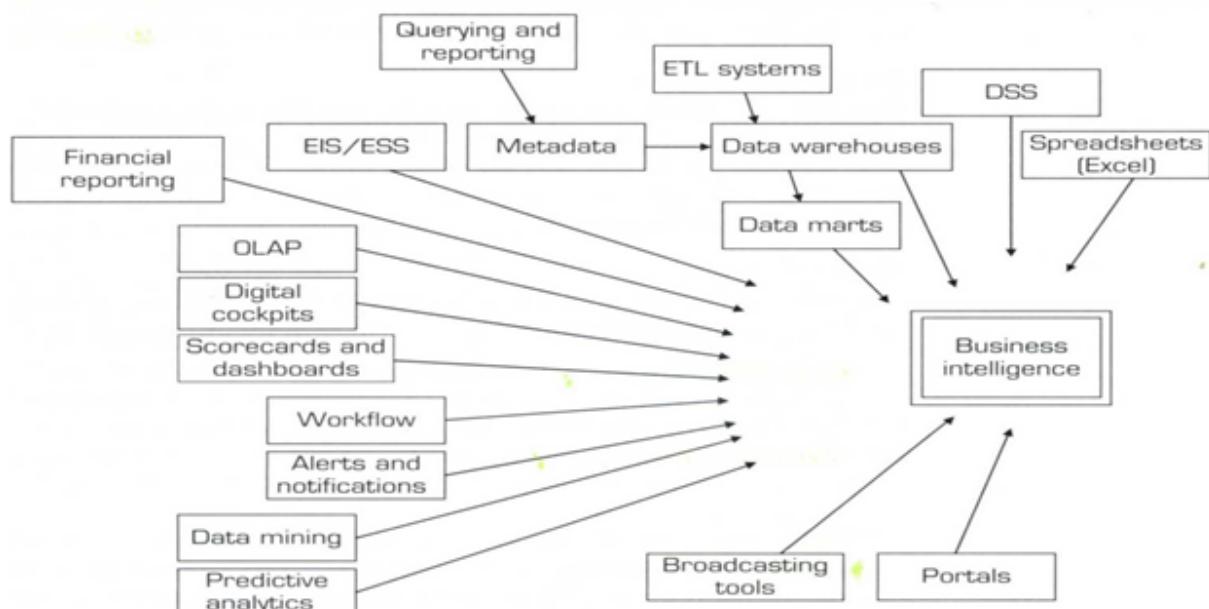
Sistemas operacionales	Sistemas Decisionales
Orientados a la transacción	Orientados a la consulta
Transacciones muy numerosas	Consultas poco numerosas
Transacciones cortas	Consultas complejas
Transacciones con pocas filas de datos involucradas	Uniones de tablas muy voluminosas que involucran a la totalidad de las filas

- **Software especializado**, que incluye motores de datos especializados, herramientas para el diseño, construcción,

alimentación y explotación del sistema. Dentro de esta categoría, cobran especial relevancia las herramientas de usuario final, que facilitan a éstos el acceso a la información por sus propios medios, de una forma muy intuitiva, con total libertad e independencia del departamento IT, sólo limitada por el propio modelo de datos, propiciando el acercamiento del sistema de información a los usuarios.

- Finalmente, **el software estadístico y de minería de datos**, para labores de investigación más complejas y avanzadas, constituye un apartado sumamente especializado dentro del concepto de business intelligence.

Son muchas las herramientas que se incluyen dentro del sistema de business intelligence. Para tener una visión completa de los mismos, reproducimos a continuación el esquema elaborado por Turban:



Sistemas de soporte a la toma de decisiones

Fuente: Intelligence system, Turban

De todas estas herramientas o componentes de la arquitectura de un sistema de business intelligence profundizaremos en las más relevantes: data warehouse, data mining y OLAP.



#### BIBLIOGRAFÍA RECOMENDADA

Turban, E. y Aronson, J. (1998) *Decision support systems and intelligent systems*. Prentice-hall international.

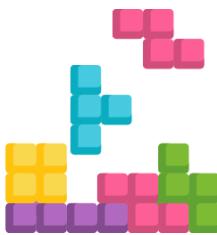
## 2.2 Ventajas del uso de BI

De forma resumida podemos decir que la implantación de un sistema de BI en una organización permitirá tomar mejores decisiones y de una forma más rápida. Pero, además, se logra:

- Un conocimiento mejor de los clientes y su comportamiento.
- La posibilidad de aumento del número de clientes.
- Obtener mayores beneficios de las ventas actuales.
- Retener clientes que consideran irse a la competencia.
- Dar significado a unos datos que aparentemente no lo tienen.
- Un proyecto de BI transforma datos en conocimiento de un modo automático, analiza grandes volúmenes de datos de un modo exhaustivo, rápido y flexible.
- Al tener una mayor información de los datos de la empresa y, por lo tanto, un mayor conocimiento del funcionamiento del negocio y del comportamiento de los clientes, podremos conseguir mayores beneficios.

Podemos clasificar las razones para su implantación en nuestra empresa de la siguiente manera:

- Motivos de negocio:
  - Competencia cada vez más agresiva y mejor preparada.
  - Necesidad de que los sistemas de información ayuden más y de forma más rápida, involucrándose en las necesidades de negocio de la empresa.
  - Clientes cada vez más exigentes, que estudian mejor la oferta de las empresas.
- Motivos tecnológicos:
  - Disponibilidad de nuevos productos software de gran complejidad.
  - Con un almacenamiento masivo de datos resulta imposible analizar por los métodos clásicos.
  - Sistemas con gran capacidad de cálculo.
  - Nuevo enfoque de los sistemas de información.



## EJEMPLO PRÁCTICO

Te han pedido que acompañes al comercial de tu empresa de servicios informáticos para realizar la venta de un sistema de business intelligence a una empresa de comida rápida.

Debes preparar un listado de beneficios para convencer al cliente de que os contrate.

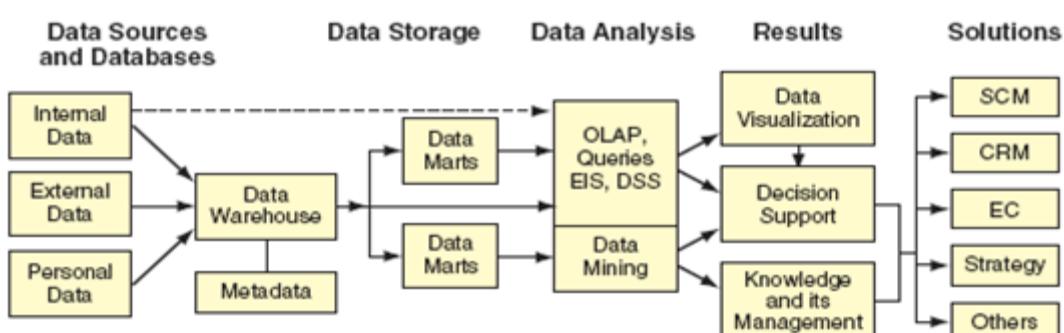
## SOLUCIÓN

Entre las ventajas que haría ver al cliente están:

- Conocer mejor a los clientes y su comportamiento, analizando los tickets de venta y encontrando relaciones entre los productos adquiridos y la hora de la venta.
- Entender el perfil de los clientes que me comprar para hacer publicidad sobre potenciales clientes con un perfil similar y aumentar las ventas.
- Analizar las horas donde se producen menos ventas para reducir los costes, por ejemplo, mediante la reducción de personal en esas horas de menor actividad.

## 2.3 El análisis de datos en BI

El proceso de análisis de datos, con las etapas que lo componen, se muestra de forma clara en el siguiente gráfico:

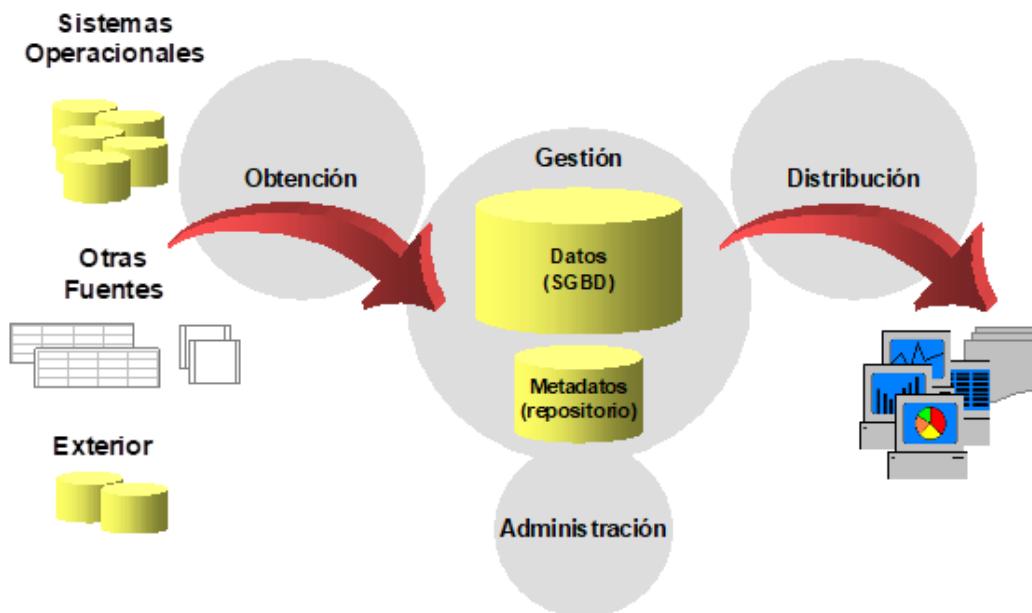


Análisis de datos empresarial

Fuente: Decision Support And Business Intelligence Systems, Turban

Es decir, a partir de unas fuentes de datos, aplicando una serie de herramientas para su almacenamiento y análisis, alcanzamos unos resultados que nos facilitan la toma de decisiones.

A grandes rasgos, cualquier solución de business intelligence gira en torno a tres componentes lógicos o fases: obtención de la información, gestión de la misma y distribución a los usuarios. La siguiente ilustración muestra la relación entre los tres componentes:



Fases en un sistema de business intelligence

Fuente: Elaboración propia

### 2.3.1 La obtención de la información

*En tu proyecto cuentas con las fuentes internas de tu empresa y con la base de datos de leads que el director comercial quiere comprar, pero crees que puedes investigar un poco y encontrar otras fuentes de datos que te permitan enriquecer tu proceso de obtención de información.*

Esta fase tiene como objetivo garantizar la realimentación automática, continua y optimizada del depósito de información, así como la calidad de esta.

Se inicia con la identificación de las fuentes de datos que van a constituir la materia prima del futuro almacén. Un porcentaje muy elevado de los datos proviene del interior de la propia compañía (aplicaciones operativas, información histórica, hojas de cálculo...) y el resto proceden del exterior, mediante adquisición a terceras compañías especializadas, o bien procedentes de algún organismo público, de la patronal del sector, etc.

Una vez extraída de las fuentes, la materia prima (datos), debe sufrir un proceso de transformación, como si de una factoría se tratara, con el objetivo de convertirla en información de calidad, eficaz y aprovechable por el nuevo sistema. El proceso de transformación incluye el filtrado de los

datos útiles, la depuración y limpieza de los mismos y su adaptación al modelo destino. En algunas ocasiones, puede incluso llegar a ser necesario desarrollar programas a medida para informar a mano sobre determinados atributos que no figuran en las fuentes. El proceso termina con la alimentación e integración de la información en el almacén.

La fase de obtención de la información es la más laboriosa de todas las que constituyen el proceso de construcción de la solución. Algunos expertos cifran hasta en un 80% el esfuerzo dedicado a esta fase respecto del total. Este porcentaje dependerá de diversos factores:

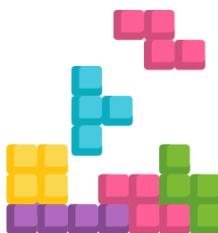
- Complejidad del modelo de datos del almacén.
- Necesidad de información exterior, importada en variedad de formatos posibles, recodificación, validación e integración de esta en el nuevo sistema.
- Fuentes internas residentes en plataformas hardware, sistemas de almacenamiento o formatos heterogéneos, incluyendo la información histórica.
- Falta de integración entre los distintos modelos operacionales, redundancias, etc.
- Falta de consistencia de los datos operacionales, información incompleta, inexacta o errónea.
- Necesidad de unificación de códigos o recodificación.

La alimentación del almacén normalmente no exige una actualización constante y sincronizada con las operaciones realizadas en cada momento, sino que se establece en términos discretos, con una carga inicial muy voluminosa que incluye toda la información histórica necesaria y refrescos periódicos (diarios, semanales, mensuales...) con las operaciones realizadas desde el último refresco. La periodicidad de éste y el detalle de la información se determinarán en función de las necesidades a cubrir por el sistema.

En esta fase deberán participar activamente los técnicos y administradores de las aplicaciones operativas, pues son ellos quienes conocen perfectamente la ubicación de los datos y sus relaciones, su forma de introducción en el sistema, su calidad, etc. Una vez identificadas las fuentes, comenzará la fase de diseño de los procesos de extracción de los datos necesarios, de los procesos de transformación, depuración, validación, adaptación de los mismos al modelo de destino, y de los procesos de alimentación e integración en el almacén.

Los aspectos más importantes por considerar en esta fase son:

- El detalle y actualidad de la información, que afecta a la granularidad de todas las estructuras, incluido el eje temporal, lo que incide directamente en la frecuencia de refresco y en el diseño del modelo destino.
- Volumetría de los datos a refrescar y plazos de tiempo necesarios para dicho refresco, para asegurar la viabilidad del proceso.
- Realizar un desarrollo a medida o utilizar software especializado.



### EJEMPLO PRÁCTICO

Has conseguido que la cadena de comida rápida contrate vuestra oferta de sistema de business intelligence. Ahora ya lanzado el proyecto, hay que definir la fase más compleja que es la de obtención de la información.

¿Qué debes tener en cuenta para que esta etapa se desarrolle correctamente?

### SOLUCIÓN

Entre otros aspectos, debo contemplar qué información externa a la compañía será necesaria. Por ejemplo, habrá que integrar información meteorológica, para conocer como la temperatura, la lluvia, etc. Influuye en la presencia de más o menos clientes en cada restaurante de la cadena.

Habrá que estudiar con detalle todas las fuentes de información de la empresa que será necesario integrar. Seguramente la cadena de restaurantes cuente con un sistema de contabilidad, obviamente tendrá una red de TPVs conectados para tomar los pedidos, probablemente tenga una base de datos de clientes para la tarjeta de fidelización, etc.

No olvidemos que hasta el 80% del esfuerzo de un proyecto como este puede originarse por las complejidades de esta fase.

Existen productos específicos para facilitar todas las tareas relativas a la obtención de la información y que pueden simplificar enormemente la labor. Estos productos se suelen denominar ETL, por sus capacidades de extracción, transformación y carga (la L proviene del inglés load). Las principales características de estos productos son:

- Interfaz gráfica de usuario con facilidades de diseño de procesos y planificación de ejecución de procesos, por fecha y hora y/o por evento.
- Importación de los catálogos de datos de las fuentes.
- Generación automática de métodos de acceso, transformación y carga.
- Información de resultados de procesos.

- Acceso a fuentes heterogéneas (bases de datos, hojas de cálculo, documentos de texto...) y carga en múltiples sistemas.
- Transformaciones y validaciones, más o menos complejas, sobre los datos y posibilidad de incorporar programas de usuario para transformaciones sofisticadas.
- Repositorio de metadatos con capacidad de intercambiar información con repositorios de otras herramientas, lo que facilita la documentación y trazabilidad de los datos.



### PARA SABER MÁS

Logstash forma parte (es la "L") de la denominada Plataforma ELK (La "E" corresponde a Elasticsearch y la "K" a Kibana). Posteriormente se han añadido otros productos a esta plataforma, como es el caso de Beats, pero comúnmente se continúa conociendo como plataforma o stack ELK.

Esta plataforma es desarrollada y evolucionada por la compañía Elastic, bajo licencia open source. Logstash puede ser utilizada como una tecnología independiente, igual que el resto de los productos de la plataforma, pero es verdad que la buena integración entre ellos ofrece claras sinergias.

Logstash facilita los procesos de transformación y carga de datos desde múltiples fuentes hacia multitud de destinos. Además, al estar basado en tuberías (pipeline) de la información, puede ejecutar distintos procesos sobre los datos e información que están siendo trasladados. Esto quiere decir que puede ser visto como una herramienta ETL muy avanzada, porque aparte de tomar orígenes convencionales en la ETL, como bases de datos SQL, también puede cargar datos desde redes sociales, software de gestión, etc. Además, el concepto de pipeline sirve para establecer etapas de procesamiento diversas, no solamente las comunes en una herramienta ETL para convertir formatos de datos, sino que otras muchos más avanzadas de manipulación de los mismos.

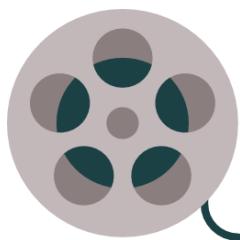
El nombre de Logstash hace referencia a que los logs de los servidores fueron el motivo principal de su creación, porque este tipo de orígenes de datos eran muy problemáticos, dado que las tecnologías utilizadas para su gestión se habían visto superadas y eran incapaces de procesarlos porque que se generaban logs más rápido de lo asumible.

Esta explosión en la generación de logs se debió a muchos motivos. Por un lado, los fabricantes de sistemas operativos, bases de datos y cualquier otro tipo de aplicaciones, comenzaron a enriquecer los logs, para con ello ofrecer información valiosa a los administradores de sistemas.

Por otra parte, comenzó un crecimiento exponencial de los dispositivos que generaban logs, en gran medida ligado al internet de las cosas. Pensemos que, desde hace años, los logs no se originan exclusivamente en los

servidores o en la electrónica de red, sino que existen cientos de millones de decodificadores de televisión en todo el mundo generando logs necesarios para monitorizar su funcionamiento y la calidad del servicio; existen miles de millones de sensores realizando todo tipo de mediciones, etc.

Finalmente, el número creciente de incidentes de seguridad informática exigieron un manejo en tiempo real de los logs de los servidores para poder detectar incidencias a tiempo.



### VIDEO DE INTERÉS

Aprende sobre el proceso ETL y su relación con la ciencia de datos con este vídeo:

<https://www.youtube.com/watch?v=x26AMwAetQ>

En general, cuanto más complejo y volátil sea el modelo de datos a implementar, más heterogéneas en formatos o plataformas sean las fuentes de información y más complejos sean los procesos de transformación, más justificados estarán estos productos dentro de la arquitectura de nuestro sistema de business intelligence.

Algunos, incluso, ofrecen herramientas para el diseño del modelo de datos del almacén de destino. De no ser así, los desarrolladores necesitarán otras herramientas para el diseño.

Algunos productos son capaces de trabajar en conjunción con propagadores de cambios. Estos tuvieron su origen en las bases de datos distribuidas, como un mecanismo de sincronización de las mismas que permitía que cualquier cambio producido en una de ellas se replicara en las demás. Su aplicación para sistemas decisionales que requieren una permanente actualización es evidente, aunque no inmediata, pues la información actualizada sobre los operacionales requiere asimismo un proceso de transformación para su adaptación al modelo destino.

Finalmente, se debe mencionar que existen en el mercado multitud de productos ETL, con una gran variedad de prestaciones y precios. Los menos sofisticados generan exclusivamente código SQL (que es un lenguaje de consulta de información en bases de datos) para la extracción y transformación, con las limitaciones propias del lenguaje, lo que en muchas ocasiones resulta claramente insuficiente. Los más potentes suponen para la organización una inversión muy fuerte que debe estar plenamente justificada.



### ENLACE DE INTERÉS

En este enlace, el comparador online de tecnologías Capterra nos ofrece su ranking de las mejores herramientas del mercado ETL actualizado al momento.

[https://www.capterra.es/directory/31099/etl/software?sort=overall\\_rating](https://www.capterra.es/directory/31099/etl/software?sort=overall_rating)

### 2.3.2 Gestión de la información

El núcleo de la arquitectura de un sistema de inteligencia de negocio lo constituye la gestión de los datos, tanto en sus aspectos de mantenimiento, como de acceso. Para esta tarea se utilizan sistemas de gestión de bases de datos. Existen dos posibles alternativas tecnológicas, complementarias y en absoluto excluyentes:

- SGBDR o sistemas de gestión de bases de datos relacionales.

Se utilizan para gestionar almacenes de información a gran escala, en los que se requiere soporte de bases de datos muy grandes. Son aplicables a los grandes sistemas de data warehouse corporativos con un elevado detalle en la información almacenada. Por supuesto, también son aplicables a sistemas más pequeños, como los denominados datamarts departamentales o específicos de un área de negocio.

- SGBDM o sistemas de gestión de bases de datos multidimensionales, también denominados servidores OLAP (o MOLAP).

Éstos utilizan estructuras muy optimizadas para el proceso OLAP, con un sistema de ficheros especializados para almacenar, comprimir y extraer conjuntos de datos con diversos niveles de sumarización, lo que proporciona un excelente rendimiento para este tipo de explotación. Estas estructuras de información se denominan cubos. El modelo de información de cada cubo recoge un solo ámbito de actividad o área de negocio, es decir, un datamart. No son eficientes para gestionar información detallada, porque los cubos producen una explosión de celdas cuando se cruzan todas las dimensiones. Es por ello por lo que su especialidad son los sistemas para el control y gestión.

Los SGBDR se han especializado tradicionalmente en sistemas transaccionales OLTP, basados en el modelo entidad-relación, cuyo objetivo es la optimización del proceso de registro y actualización de la información, donde se persigue ante todo evitar la redundancia de los datos por razones de integridad y de eficiencia al almacenarlos.

En este tipo de modelos es frecuente que existan varios caminos para resolver una misma consulta, con tiempos de respuesta diferentes para cada uno de ellos. En los últimos años los fabricantes de SGBDR han realizado un gran esfuerzo en mejorar sus prestaciones en procesos de consulta. Entre las nuevas prestaciones, cabe destacar:

- Optimizador de consultas basado en costes y reescritura automática de consultas para optimizarlas y acelerarlas.
- Distintas técnicas de indexación para acelerar los accesos por columnas a los datos.
- Partición automática de tablas voluminosas por distintos criterios y almacenamiento en diferentes dispositivos, para acelerar los accesos y aumentar la disponibilidad.
- Paralelismo, tanto en procesos de carga masiva, como en consultas (cada consulta se desmenuza en pasos más pequeños que se resuelven en paralelo aumentando la velocidad global del sistema).

Los SGBDR son más versátiles y, en general, bastante más económicos que los servidores OLAP. Estos sólo se justifican si la organización tiene necesidades de explotación OLAP de la información, pues ésta es su verdadera virtud. La utilización de ambas tecnologías es posible e incluso recomendable: se podría disponer de un SGBDR para gestionar el data warehouse y la información de mayor nivel de detalle, y de un servidor OLAP para gestionar los datos de las aplicaciones OLAP.

No obstante, también existe la alternativa de utilizar un SGBDR como motor de datos para aplicaciones OLAP. La solución, aunque es posible, es más compleja y requiere un mayor esfuerzo para su construcción y administración. En este caso, es necesario utilizar técnicas de modelización multidimensional para obtener buenos tiempos de respuesta.

La modelización multidimensional persigue la optimización de las consultas. Los accesos a la información están claramente definidos y optimizados. También es habitual almacenar información precalculada que puede deducirse por el contenido de otros elementos, todo en aras de reducir el número de accesos necesarios para obtener distintos valores.

No debe preocupar la redundancia de información que propicia el modelo multidimensional, ni por el aumento de la volumetría (el precio del

almacenamiento se ha reducido considerablemente y seguirá haciéndolo) ni por la penalización de los procesos de actualización (la explotación del modelo es en modo consulta).

Por último, debemos realizar una breve reflexión acerca de la problemática de la disponibilidad de la información y la capacidad de las líneas de comunicación remota, máxime teniendo en cuenta el coste de éstas. Las organizaciones geográficamente diseminadas que presentan un alto grado de dispersión deben considerar muy seriamente la utilización de bases de datos distribuidas como una alternativa razonable, sobre todo cuando el tamaño de la organización y el número de usuarios así lo aconseje.

Esta solución acerca los datos allá donde se necesitan, de forma que aumenta la disponibilidad del sistema y minimiza el tráfico de información, reduciendo las necesidades de comunicación remota. Como contrapartida, hay que señalar que la solución es mucho más compleja de construir y administrar, puesto que debe realizarse el análisis de qué datos se distribuyen y además construirse los procesos de sincronización de bases de datos distribuidas, tarea nada sencilla.

Además de la gestión de los datos, el sistema debe proporcionar una completa gestión de metadatos, tanto técnicos como de negocio. Los metadatos son datos de referencia sobre los objetos manejados por el sistema. Decimos que es importante el uso y gestión de metadatos porque en sistemas complejos, nos ayudarán a mantenerlo, administrarlo y explotarlo.

Los metadatos técnicos se refieren a las descripciones de los objetos almacenados en la base de datos (su nombre, descripción, origen, transformaciones...).

Los metadatos de negocio actúan como una capa de abstracción, que relaciona conceptos de negocio con los datos físicos almacenados. Cuando un usuario necesita acceder a todo o una parte del modelo de datos, mediante una herramienta de usuario final de consulta o para generar un informe, lo hará utilizando los metadatos de negocio, que le ocultan la complejidad del modelo, y actúan como una especie de plano, ayudándole a navegar por la información almacenada y comprender su significado en términos de negocio.

Es frecuente que cada una de las herramientas utilizada para la construcción y explotación del sistema disponga de su propio repositorio de metadatos. El sistema de gestión de metadatos debe permitir su integración y mantenimiento en un único repositorio. A la hora de configurar la solución, se tendrá en cuenta que un criterio especialmente importante en

la selección de las herramientas a utilizar sea su integración y capacidad de compartir metadatos entre ellas, porque facilitará enormemente la construcción, permitiendo su reutilización por todos los componentes de la solución.

Finalmente, la gestión de la información también recoge el concepto de administración, el cual se convierte en una cuestión clave cuando aumentan los volúmenes de datos y usuarios del sistema. La solución debe contar con herramientas y utilidades de administración, con el objetivo de optimizar la gestión de los datos y aumentar los rendimientos.

Las herramientas de administración se acompañan con el gestor de la base de datos, si bien no todas tienen las mismas prestaciones. Además de permitir una completa monitorización y administración del sistema a través de una sencilla pero potente interfaz gráfica (todas o casi todas lo hacen), sería muy interesante que las herramientas tuvieran capacidades de sugerir la creación de índices y tablas sumarizadas, en base a las estadísticas de uso, así como capacidades para la creación y el refresco automático de las tablas sumarizadas cuando así lo decida el administrador. También que dispusiera de completos asistentes de administración para facilitar la labor al administrador.

### **2.3.3 Distribución de la información**

Los usuarios son los que deberán decidir las necesidades de información, lo que va a determinar tanto el diseño del modelo de datos como de los procesos de obtención de información, y también va a condicionar la estrategia para su gestión. La fase de distribución constituye la exteriorización de la totalidad del sistema y donde se van a apreciar las ventajas de disponer de una plataforma orientada a la decisión. De cómo éstas sean percibidas por los usuarios, depende en gran medida el éxito de todo el esfuerzo realizado. Para garantizar el éxito, debe comprometerse a todos los departamentos y usuarios afectados en su definición precisa.

Como ya se ha mencionado anteriormente, estos sistemas tienen distintas aplicaciones, y existen diferentes alternativas tecnológicas en el mercado. El proceso completo de construcción, así como las herramientas y tecnología seleccionadas, deberán ajustarse a las necesidades de explotación manifestadas por los usuarios. En esta fase se utilizarán dos tipos de software: de acceso a la información y de suministro de información. No son excluyentes, sino que cada uno tiene su utilidad y aplicación. La siguiente tabla analiza sus aplicaciones y características:

Acceso a la información	Suministro de información
Se aplican para el proceso OLAP, EIS, consultas ad-hoc y minería de datos.	Se aplican para el reporting empresarial y departamental.
Se ponen a disposición de determinados usuarios.	Son corporativas, es decir, se ponen a disposición de toda la organización.
Dejan en manos de los usuarios la responsabilidad de explotar la información según las necesidades de cada uno y de cada momento (libertad e indeterminación).	Ofrecen un enfoque planificado y gestionado por el sistema, lo que supone un diseño y construcción previos (planificación y determinación).
Requieren de una adecuada interactividad sistema/usuario, al considerar la plataforma como un elemento pasivo de la solución, que responde a las solicitudes del usuario.	Establecen una plataforma de suministro de información activa, que genera y distribuye por sí misma la información que precisan los usuarios.

Existe un aspecto tecnológico que afecta a la estrategia de despliegue de la solución a lo largo de toda la organización. En organizaciones pequeñas y concentradas, no supone mayor problema la instalación de este software y su sustitución por nuevas versiones. En el otro extremo, las organizaciones de gran tamaño con un alto grado de dispersión geográfica deberán plantearse muy seriamente la problemática de distribución del software.

Por este motivo, la alternativa que ha ganado terreno a la tradicional arquitectura cliente/servidor es la utilización de clientes ligeros que sólo necesitan disponer de un navegador web para poder operar con el sistema. Esta alternativa utiliza la tecnología web como estrategia de despliegue de la solución.

La actualización del software sólo debe realizarse en el servidor web, y se descargará automáticamente en las estaciones de trabajo. Muchas soluciones están disponibles en la doble modalidad cliente/servidor y cliente ligero. Como contrapartida, no todos los fabricantes disponen de soluciones en cliente ligero y algunas que sí están disponibles pierden funcionalidad respecto a su versión cliente/servidor.

## 2.4 Data warehouse

Desde hace bastantes años, cuando se popularizó el término data warehouse, éste se concibe como un gran almacén de datos corporativo, donde se integra toda la información de gestión de la compañía, procedente de múltiples sistemas operacionales.

Este almacén debe ser suficiente para proporcionar cualquier información de gestión que le sea solicitada, a cualquier nivel de detalle, de cualquier periodo histórico y de cualquier área de negocio de la compañía.

No debe confundirse el data warehouse con una base de datos corporativa contra la que las aplicaciones realizan procesos transaccionales, ni con una copia centralizada de los datos generados por dichas aplicaciones. El concepto que define un data warehouse es mucho más que un gestor de bases de datos o el software especializado.

Numerosos han sido los proyectos de data warehouse que han fracasado debido a la gran ambición de los objetivos planteados o, dicho de otro modo, por la falta de concreción de los mismos. Por otro lado, el esfuerzo para la construcción del data warehouse tarda mucho en dar sus frutos, y estos no son los esperados en numerosas ocasiones. Además, se produce una falta de identificación con el proyecto por parte de los distintos gestores, debido fundamentalmente al temor de que la solución no resulte idónea para el ámbito de su actividad.

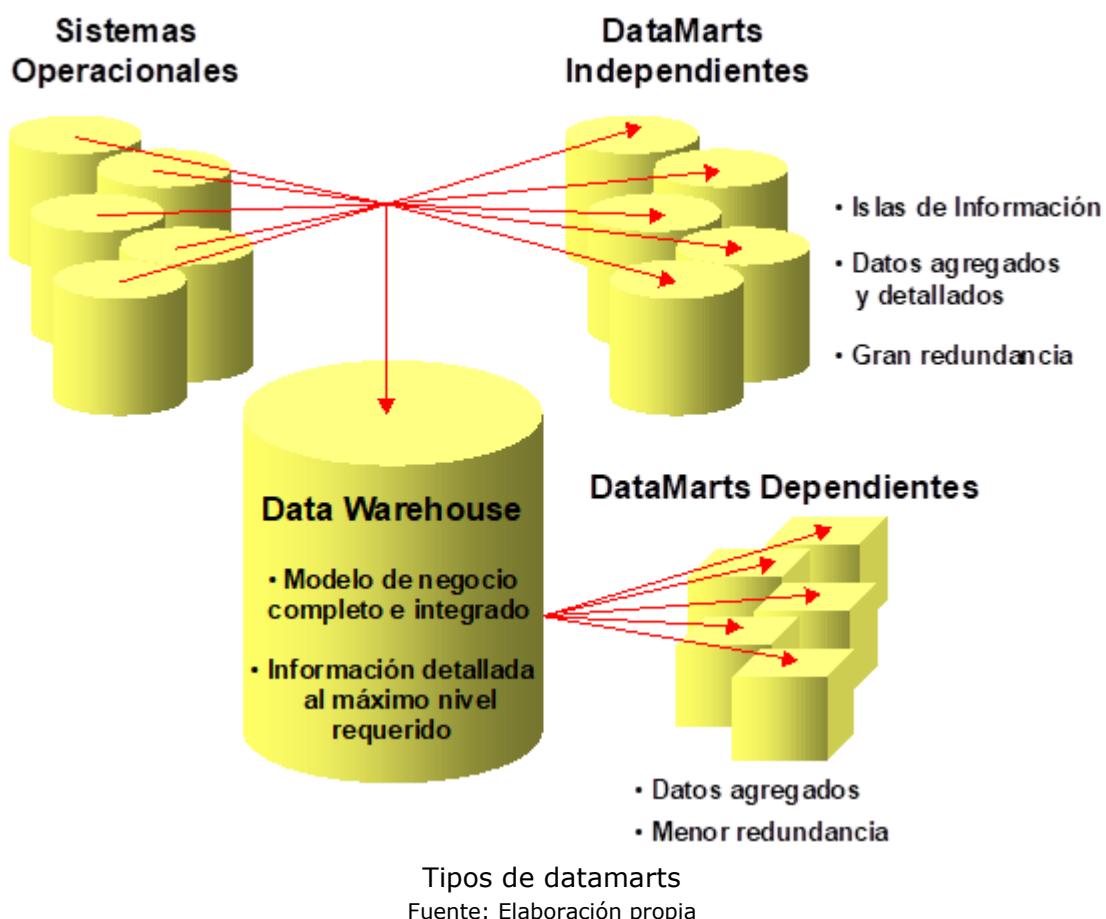
Tras estas experiencias, la tendencia actual del mercado podría resumirse en la frase "Piensa globalmente y actúa particularmente", es decir, no se abandona la idea de disponer en el medio/largo plazo de un gran almacén unificado y fuente de todo el conocimiento empresarial, pero su construcción se aborda por etapas. Cada una de estas etapas se corresponde con cada uno de los diferentes departamentos, o más correctamente, áreas de negocio o de actividad que van a constituir el data warehouse. Esta metodología presenta la gran ventaja de ofrecer resultados en el corto/medio plazo, lo que permite a la organización seguir apostando por el proceso, además de aumentar la identificación de los gestores en cada ámbito de actividad.

Cuando se implementa el modelo de negocio de un ámbito de actividad, se crea lo que se denomina un datamart. Éste tiene las mismas características del data warehouse (comprensión, calidad, orientación al análisis del negocio, riqueza) pero el alcance funcional del modelo se limita a un departamento o área de negocio.

A medida que se van implementando nuevos datamarts, se recomienda que vayan integrándose en un único y gran repositorio, de tal forma que al final

del proceso se disponga del data warehouse corporativo. Esta metodología da lugar a lo que se denominan datamarts dependientes, que se alimentan no ya desde los sistemas operacionales, sino desde el propio data warehouse. En este se mantiene toda la información de gestión de la compañía, detallada al máximo nivel requerido. Mediante un proceso de distribución, se generarán los datamart departamentales.

También se puede optar por mantener datamarts independientes, sin ningún tipo de integración entre sí, que se alimentan directamente desde los sistemas operacionales en su mayor parte. Esta metodología permite una rápida implementación y no exige ningún esfuerzo adicional de integración. La siguiente ilustración muestra de forma gráfica el resultado que se obtiene en ambas implementaciones:



Los datamarts independientes permiten una implementación más rápida y no requieren ningún esfuerzo de integración. Sin embargo, se muestran menos eficientes y más difíciles de mantener en el largo plazo. La gran virtud del data warehouse unificado es la alta reutilización que proporciona, es decir, disponer de una parte importante de cualquier nuevo sistema ya construido y listo para su utilización.

Cualquier departamento que necesite tomar decisiones sobre un ámbito del negocio solicitará la construcción de un datamart específico. El data warehouse permitirá la obtención de esos datamart específicos mediante un sencillo proceso de distribución.



#### SABÍAS QUE...

A Bill Inmon se le considera universalmente como el padre del data warehouse. Su primera definición del mismo fue: «un almacén de datos es la recopilación de datos orientados e integrados y dependientes del tiempo que son la base de la toma de decisiones en la gestión»

Inmon nació en 1945 y todavía en la actualidad lidera su empresa Forest Rim Technology orientada actualmente al mercado de big data y análisis.

Las características que Inmon definió en el año 2005 para un data warehouse en relación al proceso de toma de decisiones fueron las siguientes:

- Los datos se estructuran por temas específicos, como pueden ser proveedores, productos o clientes, y contienen solo la información que pueda ser pertinente para tomar una decisión. La orientación a temas facilita a los usuarios determinar no sólo cómo se desarrollan sus negocios sino también el porqué.
- Un data warehouse se diferencia de una base de datos convencional en que la mayoría de las bases de datos tienen una orientación al producto, mientras que un data warehouse proporciona una visión más completa de la organización, al estar orientado a los temas o líneas de negocio estratégicos de la empresa.
- La integración está estrechamente relacionada con la orientación a temas. Los almacenes de datos deben situar datos de diferentes fuentes en un formato consistente. Para conseguir este objetivo, deben considerar conflictos y discrepancias de nombre entre unidades de medida.
- El tiempo es una dimensión importante que deben soportar todos los data warehouse, ya que los datos para análisis de múltiples fuentes contienen puntos temporales que nos permiten detectar tendencias o desviaciones.

## 2.4.1 Profundizando en la arquitectura del data warehouse

Como vamos observando, el concepto de business intelligence envuelve distintos términos y componentes clave que deben ser estudiados, entendidos, cuantificados y analizados en el momento de plantear una estrategia de BI para tu propia empresa o para un cliente.

De igual forma, como también se ha reseñado, la tecnología juega un papel determinante en el crecimiento y/o evolución de este modelo. Como sabemos, en la era digital la generación de datos a nivel corporativo puede ser un elemento útil para generar valor.

Dado que este aumento de la importancia de los datos parece que no va a detenerse por ahora, para comprender por donde puede ir el futuro, veamos tres reglas básicas de la evolución de la tecnología:

- **Ley de Moore:** Esta indica básicamente que el poder de procesamiento de un computador se duplica cada 18 meses.
- **Ley de Butter:** En esta ley se establece que las capacidades de comunicación duplican la velocidad cada 9 meses.
- **Ley de Kryder:** Por esta ley se establece que la capacidad de almacenamiento por centímetro cuadrado se duplica cada 13 meses.

En lo que tiene que ver con la gestión de datos, es especialmente importante lo que establece la ley Kryder.

El Ingeniero Mark Kryder formuló esta observación hacia principios del año 2000, pero a lo largo del tiempo se ha detectado que esa tendencia realmente se ha ralentizado, puesto la capacidad se duplica actualmente cada 15 o 17 meses. Pero, a pesar de ese frenazo, claramente esto sigue siendo mucho más rápido que la primera ley mencionada.

Para verlo en términos prácticos, seguramente todos nosotros recordemos en nuestra casa o colegio el primer ordenador con el cual tuvimos contacto. Seguramente este ordenador, fuese de sobremesa o portátil tenía una capacidad de almacenamiento inferior al que hoy tenemos en un pendrive USB o incluso en nuestro smartphone.

Estas leyes plasman de forma conceptual o teórica como la tecnología evoluciona traduciéndose en mejoras de rendimientos en los dispositivos y disminución de los costes de adquisición. Es así como se introdujo y potenció el concepto de data warehouse.

También ha potenciado la asimilación de los data warehouse un fenómeno que seguramente todos hemos sufrido y que se ha denominado como "Excel caos". Este caos provocado por el uso intensivo de hojas de cálculo en las empresas nos lleva a que existan diversas fuentes y ficheros de información separados con múltiples nombres o versiones incontrolables. De ahí surge la

necesidad de centralizar la información, controlarla y permitir un acceso fácil.

Prueba de esto es el uso de gestores de bases de datos especializados como por ejemplo Oracle o MongoDB, el desarrollo de metodologías que buscan formar y capacitar a los involucrados en el gobierno de datos y la arquitectura para afrontar el reto del crecimiento exponencial de acuerdo a las tres dimensiones que vimos: volumen, velocidad y variedad y más recientemente el internet de las cosas han facilitado la migración a la nube teniendo así fácil acceso a soluciones tipo SaaS (Software as a Service), lo que da ventajas significativas a la hora de realizar integraciones o implementaciones rápidas de nuevas estructuras de datos y sistemas.

Por este motivo, dado que el data warehouse es la piedra angular de una arquitectura de BI, vamos a profundizar también en la arquitectura del propio data warehouse, estableciendo los conceptos base, definiciones comunes y prácticas recomendadas para diseñar la arquitectura de un data warehouse y como este soporta el modelo de inteligencia de negocio.

Los cimientos del BI se establecen sobre el control eficiente y efectivo de los datos. Por dicho motivo es necesario crear un repositorio de datos que permita centralizar y controlar la información, estableciendo protocolos, procesos y políticas para su tratamiento.

Entonces diremos que un data warehouse es un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, que tienen por objetivo dar soporte a la toma de decisiones.

Algunas de las características y atributos que debe tener el data warehouse son:

- Estable
- Coherente
- Fiable
- Información histórica.

Por tanto, las bases de datos relacionales resultan muy útiles como soporte técnico para almacenar y visualizar las estructuras de datos. En términos resumidos, las características que debe tener el data warehouse acorde a la definición de Inmon son:

- Orientado a un tema: La colección de datos se enfoca a un tema: atención al cliente, compras, ventas etc.
- Integrado: Los datos son variados, tienen distintas formas y orígenes, pero siempre siendo consistentes.
- No volátil: Los datos únicamente son para lectura de los usuarios.

- Variable en el tiempo: Se realizan capturas de información en períodos de tiempo concretos y estableciendo una cronología.

Aunque un enfoque típico para gestionar los datos son las bases de datos relacionales, estas no son la única forma de abordar la construcción, pues existen -por ejemplo- bases de datos asociativas u orientadas a columnas.

En cualquier caso, deberá realizarse un análisis de cuál es el modelo de base de datos más apropiada, basándose en la necesidad final a cubrir, los objetivos estratégicos y la capacidad de inversión dado que, aunque existen recursos open source o de código abierto, algunos escenarios nos obligarán a recursos de pago -en forma de licencia o como servicio- que ofrezcan otro tipo de ventajas que podamos necesitar.

Para comprender de mejor forma y tener un mejor contexto para entender la arquitectura del data warehouse dentro del business intelligence, definamos los siguientes términos:

- **Data warehousing:** Es un proceso para extraer y filtrar datos de las operaciones y transacciones de la organización y sus sistemas, para transformarlos, integrarlos y almacenarlos.
- **Datamart:** Es un subconjunto de datos dentro del data warehouse, el cual tiene por objetivo específico facilitar el realizar un determinado análisis o función para un grupo de usuarios determinado.
- **Operational data store (ODS):** Es un tipo de almacén de datos el cual proporciona únicamente los últimos valores o los más recientes excluyendo el histórico.
- **Staging area:** Es un sistema que permanece entre las fuentes de datos y el data warehouse con el fin de facilitar la extracción de datos, mejorar la calidad de estos y ser usado como un caché para luego realizar el data warehousing.
- **Procesos ETL:** Es una tecnología de integración de datos basado en tres pasos, extraer, transformar y cargar, la cual se usa tradicionalmente en la integración y diseño del data warehouse.
- **Metadatos:** Son datos que aportan la interpretación de una instancia u ofrecen la información para administrar, descubrir y analizar diferentes solicitudes.

Una base de datos operacional suele respetar la forma relacional en su diseño, pero para un data warehouse este aspecto puede variar, pues la idea es que la información esté desnormalizada, con el fin que se realicen las consultas de forma óptima.

Por tanto, como ya se ha mencionado con anterioridad, una aproximación inicial siempre debe girar en torno a identificar los procesos de negocio

involucrados, sus objetivos, esquemas típicos de evaluación y sus medidas asociadas. Consideremos, por ello, siempre los siguientes aspectos:

- **Tablas de hecho:** Es la representación gráfica de las relaciones en un proceso de la organización. Por ejemplo: Un pedido en un comercio puede identificarse como un proceso, por lo cual se debe considerar una tabla del hecho pedidos.
- **Dimensiones:** Es la representación en el data warehouse de una visión del negocio, a través de un proceso de negocio. Siguiendo con el ejemplo de un pedido del cliente y la fecha de creación del pedido, esto se considera como una vista para este proceso, lo que resultará en la recopilación de todos los pedidos de un cliente.
- **Métricas:** Estos son propiamente las medidas del proceso de negocio involucrado de forma cuantificable como, por ejemplo, valor del pedido.

Se encuentran principalmente dos esquemas de representación de relaciones de datos y su estructuración:

- **Esquema en estrella:** consiste en estructurar la información en procesos, vistas y métricas recordando a una estrella (de ahí el nombre).
  - A nivel de diseño, consiste en una tabla de hechos (lo que en los libros y manuales técnicos encontraremos normalmente como fact table) en el centro, para el hecho objeto de análisis y una o varias tablas de dimensión por cada punto de vista de análisis que participa de la descripción de ese hecho.
  - En la tabla de hecho encontramos los atributos destinados a medir (cuantificar): sus métricas. La tabla de hechos sólo presenta uniones con dimensiones.
- **Esquema en copo de nieve:** es un esquema de representación derivado del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas y aparecen nuevas uniones. Es posible distinguir dos tipos de esquemas en copo de nieve.



#### BIBLIOGRAFÍA RECOMENDADA

Curto, J. (2012) *Introducción al Business Intelligence*. Barcelona: Editorial OUC.

Así mismo veamos una aproximación a los tipos de tablas de hecho y posibles métricas y tipos de dimensiones. Las tablas de hecho son una

representación del proceso de negocio, que como tabla permite guardar dos tipos de atributos:

- Medidas del proceso
- Claves foráneas.

Los cuatro tipos de tablas de hecho que pueden emplearse y sobre las cuales trabajaremos con más detalle en las unidades dedicadas específicamente al data warehouse y daramarts son:

- Transaction fact table.
- Factless fact tables / Coverage table
- Periodic snapshot fact table
- Accumulating snapshot fact table

Para el caso de las dimensiones, estas recogen los puntos definidos en el proceso donde se debe realizar algún tipo de análisis de un evento o hecho ocurrido. Desde una perspectiva de la gestión histórica de los datos existen seis dimensiones principales:

- SCD2 tipo 0: No se considera la gestión del cambio de datos históricos.
- SCD2 tipo 1: No se guardan históricos, la nueva información se sobre escribe.
- SCD2 tipo 2: Toda la información histórica se guarda en el data warehouse.
- SCD2 tipo 3: Toda la información histórica se guarda en el data warehouse. Se crean columnas con nuevos valores y los actuales se reemplazan.
- SCD2 tipo 4: Es lo que se conoce como tablas históricas, hay tablas para datos nuevos y para datos antiguos.
- SCD2 tipo 6/Hibrida: Combina las aproximaciones  $1+2+3=6$ . Esto quiere decir que se considera una dimensión de tipo 1 y añade un par de columnas adicionales que indican el rango temporal de validez de una de las columnas de la tabla.

Para las métricas estableceremos distintos tipos de medidas, las cuales se basan en el tipo de información que se busca recopilar, asociado a la funcionalidad esperada de acuerdo con la visión del proceso de negocio.

- Medidas de realización de actividad
- Medidas de resultado de una actividad

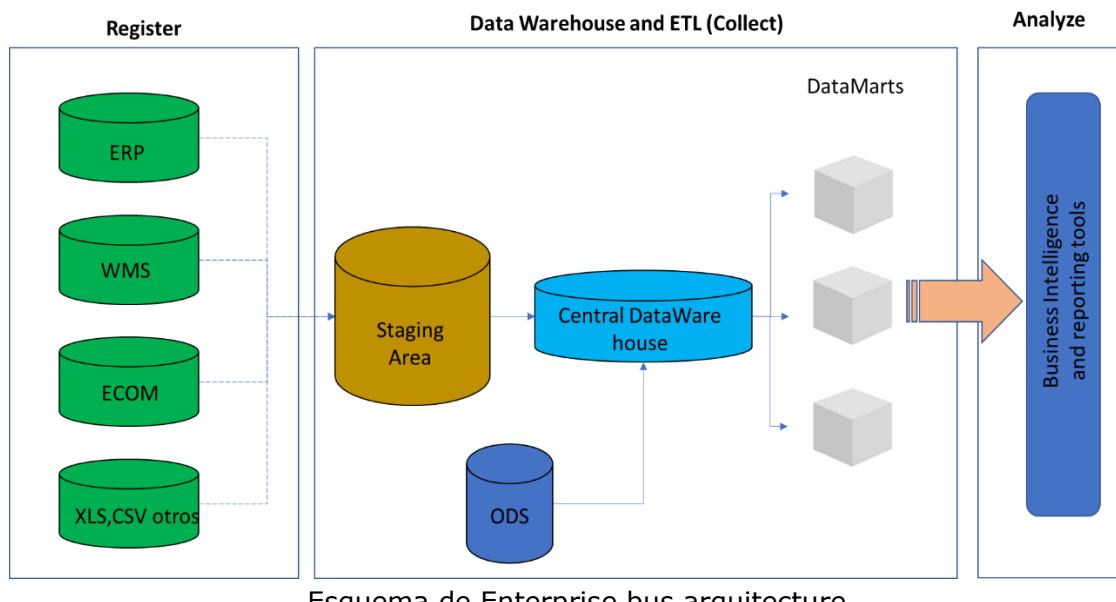
Para estos tipos de medida se crean o definen los indicadores claves de desempeño, comúnmente llamados KPI (Key Performance Indicator), que formarán parte de un cuadro de mando integral con la visión de la compañía y que buscan medir el desempeño o rendimiento del proceso.

Asimismo, también se incluyen los KGI (Key Goal Indicator), enfocados a la consecución de metas del proceso de negocio. En todo caso, estas medidas, tablas de hecho y dimensiones deben ser siempre observadas y diseñadas con el fin de lograr los objetivos estratégicos de la organización basados en datos.

#### **2.4.2 Variaciones en la arquitectura de data warehouse**

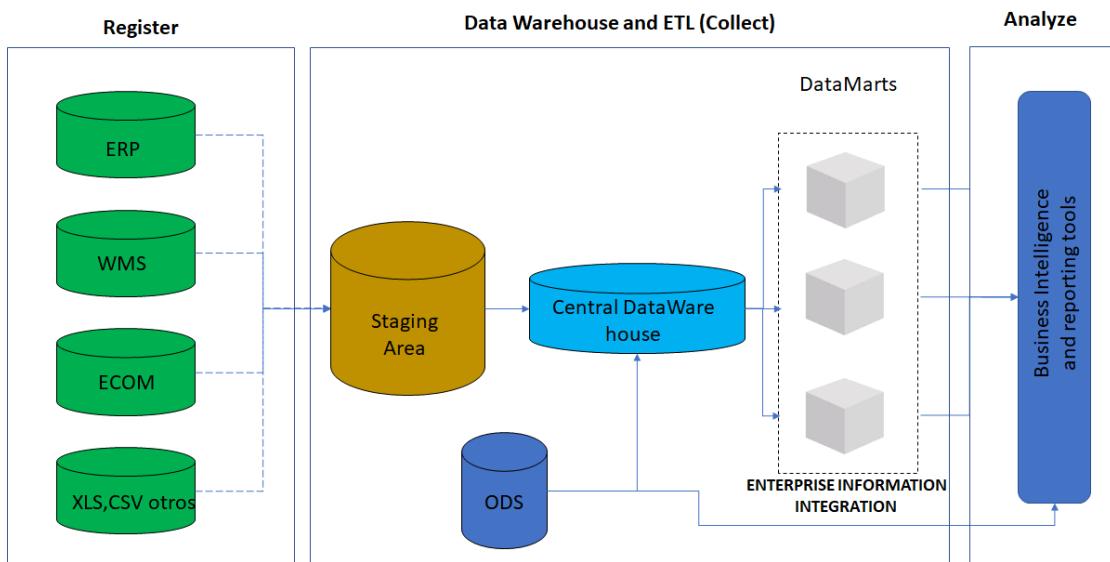
Entre las modalidades de abordar la arquitectura del data warehouse nos encontramos con:

- Enterprise bus architecture (o data warehouse virtual/federado), también conocido como MD (multidimensional architecture): consiste en una arquitectura basada en datamarts independientes federados, que pueden hacer uso de una staging area en el caso de ser necesario. Federados quiere decir que se hace uso de una herramienta EII (Enterprise Information Integration) para realizar las consultas como si se tratara de un único data warehouse. Puede existir en el caso de ser necesario un ODS.



Fuente: Chapell, D. (2004). Enterprise Service Bus: Theory In Practice

- Corporate information factory (o enterprise data warehouse): consiste en una arquitectura en la que existe un data warehouse corporativo y unos datamarts (o incluso cubos OLAP) dependientes de este. El acceso a datos se realiza a los datamarts o a la ODS en caso de existir, pero nunca al propio data warehouse. Puede existir - en el caso de ser necesaria- una staging área.

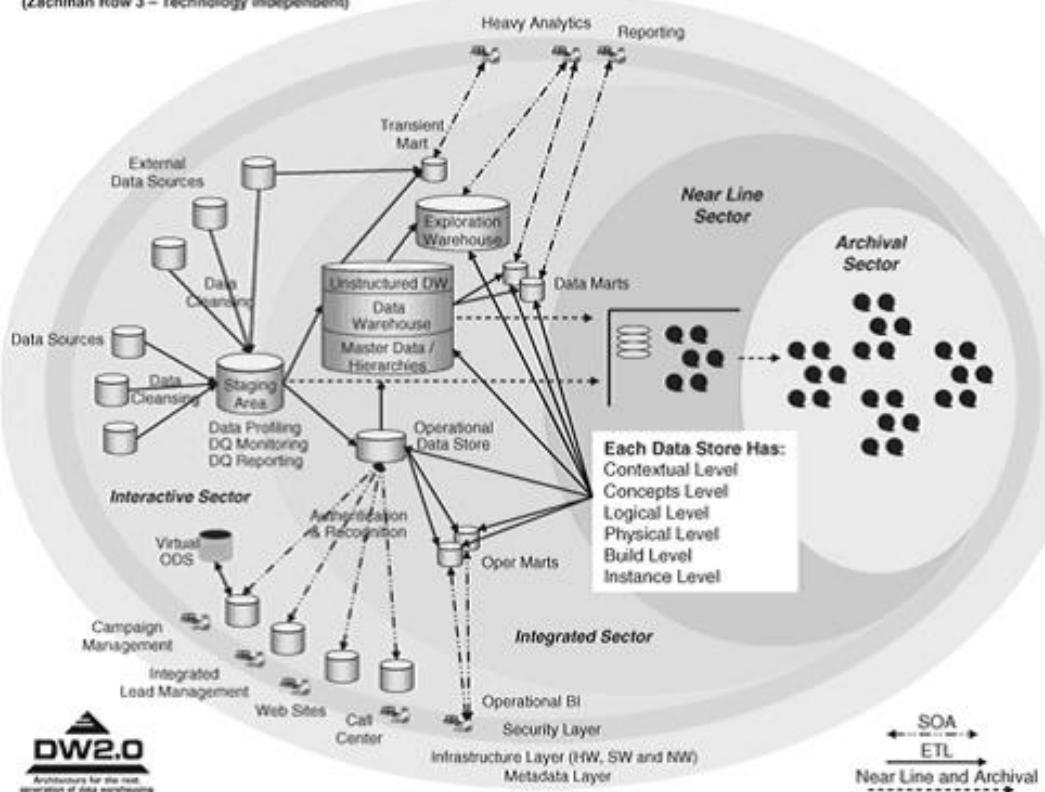


Esquema de Corporate information factory

Fuente: Chapell, D. (2004). Enterprise Service Bus: Theory In Practice

- Enterprise Data Warehouse 2.0: consiste en la revisión de la metodología de Bill Inmon para incluir toda la experiencia de los últimos años. El punto diferencial es que se separa la información por la edad de esta y se clasifica por su uso. Se caracteriza por completar tanto la inclusión de información estructurada como no estructurada y por focalizarse en el objetivo de responder a todas las necesidades actuales de negocio. El siguiente gráfico representa una arquitectura completa.

STRATEGIC Logical Architecture Schematic  
(Zachman Row 3 – Technology Independent)



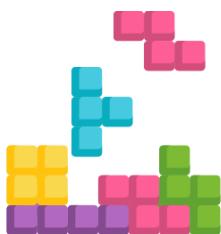
### Esquema de Enterprise Data Warehouse 2.0

Fuente: Chapell, D. (2004). Enterprise Service Bus: Theory In Practice

Esta arquitectura permite distinguir varios tipos de entornos en función de la información y su uso. Esta metodología del DW 2.0 aún sigue desarrollándose, estando en un proceso importante de despliegue y suele combinarse con data vault. Un modelo que se basa en tres tipos de entidades:

- Hubs: contiene indicadores claves
- Links: contiene todas las relaciones
- Satélites: contiene las descripciones.

Este diseño obedece a la búsqueda de la máxima flexibilidad y optimización del data warehouse, pensando en adaptaciones, integraciones y evoluciones futuras del negocio para facilitar su escalabilidad.



### EJEMPLO PRÁCTICO

Un gran banco ha entrado en una situación muy preocupante debido a la crisis económica que está pasando el país donde lleva a cabo sus negocios. Los directivos del banco han valorado que la mejor forma de enfrentarla es mejorar sus procesos de aprobación de préstamos. Es decir, ser muy cuidadosos y solo prestar dinero a las empresas que de verdad puedan devolverlo.

¿Qué enfoque sería adecuado desde el punto de vista de crear un data warehouse?

### SOLUCIÓN

Dada la situación, hay que actuar con urgencia, por lo que plantearse la creación de un gran data warehouse integrado no es la mejor opción.

Probablemente la mejor opción es crear un data mart inicial independiente, solo para el departamento de riesgos y préstamos del banco, de forma que los analistas cuenten con toda la información cargada desde los sistemas operacionales. Es decir, un analista podría ver los cobros y pagos de cada cliente del banco; podría analizar el saldo medio en las cuentas, las inversiones, el uso que ha hecho de los préstamos y créditos anteriormente concedidos, etc.

De esta forma, conseguimos una rápida implementación, sin esfuerzo adicional de integración.

Ya en el futuro, cuando la situación mejore, se podría plantear el crear otros data marts para marketing, recursos humanos, etc. E integrarlos todos en un data warehouse para mejorar las operaciones globales del banco.

## 2.5 Data mining (minería de datos)

La minería de datos es una herramienta típica que forma parte de la arquitectura de los sistemas de BI y de DSS, consistente en la búsqueda y obtención de conocimiento del comportamiento de negocio a partir de la información disponible en la organización y de las relaciones entre los datos.

La minería de datos se basa en un conjunto de técnicas de selección, exploración y modelización de grandes volúmenes de información para llevar a cabo esta labor. Para acelerar todo el proceso, normalmente se

utiliza un software especializado en las citadas técnicas. El proceso de minería de datos es complejo y debe abordarse de forma conjunta entre los expertos del negocio y los expertos en las diferentes técnicas aplicables, generalmente de tipo estadístico.

El proceso de minería de datos consiste en un conjunto de técnicas de análisis que, aplicadas sobre los datos disponibles en una organización, permite obtener conocimiento del comportamiento de negocio. Estas técnicas, de tipo estadístico e inteligencia artificial, necesitan analizar numerosas observaciones para obtener resultados fiables, es decir, la reducción del posible error cometido se consigue ampliando la muestra.

El software cubre todo el proceso, desde la selección de los datos (muestreo) y la exploración de los mismos, hasta la construcción de complejos modelos predictivos.

Si el proceso no se realiza adecuadamente, se pueden obtener conclusiones obvias, absurdas o incorrectas. Es un proceso cíclico, por cuanto que se realimenta con los resultados para mejorar los modelos, es decir, obtenido un primer modelo que intenta reproducir una determinada situación, se analizan los resultados que proporciona, se ajustan las variables y criterios empleados y se obtiene un nuevo modelo más acorde a la realidad que trata de modelizar.

Algunas de las aplicaciones de la minería de datos entran en una de las siguientes categorías:

- Gestión del mercado: Enfocada a las campañas de marketing y promocionales. Trabaja sobre bases de datos corporativas que almacenan preferencias de los clientes e información pública sobre la demografía del cliente y sobre su estilo de vida. Los algoritmos determinan segmentos de clientes sobre los que centrar el esfuerzo de marketing.

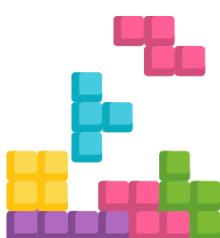
Parece que gran parte del proyecto que te ha encargado tu director comercial va a ir por estos derroteros.

- Gestión de riesgos: Se ocupa de analizar los riesgos del negocio originados en la amenaza competitiva, la insuficiente calidad de los productos y el abandono de los clientes hacia los competidores. Los algoritmos construyen modelos de clientes vulnerables susceptibles de abandonar la compañía. Por ejemplo, en el sector de seguros, estas aplicaciones están orientadas a predecir las pérdidas por daños sobre un conjunto de suscriptores de pólizas o el riesgo total de pérdidas.

- Gestión del fraude: Enfocadas a compañías donde se procesan una enorme cantidad de transacciones que las hacen más vulnerables al fraude. Los algoritmos detectan transacciones fuera de rango según la actividad de una persona, reteniéndolas para su posterior comprobación.

Algunas de las operaciones que se pueden realizar con un proceso de minería de datos son:

- Construcción de modelos predictivos: En base a observaciones previas, se construye un modelo sobre un cierto concepto que utilizamos después para predecir el futuro comportamiento de nuevos elementos. Por ejemplo, se puede construir el modelo de cliente leal de una compañía para determinar el riesgo de pérdida de determinados clientes.
- Segmentación de bases de datos (clustering): Se divide la base de datos en segmentos que contienen registros que comparten determinadas propiedades definidas por el valor de sus atributos. Los algoritmos descubren qué atributos definen la partición. Por ejemplo, "varón, urbano, soltero y altas rentas" puede ser un segmento de consumidores interesante para una determinada campaña de marketing.
- Análisis de relaciones: Consiste en descubrir asociaciones entre registros de una base de datos, y por analogía, entre los objetos básicos del negocio. Por ejemplo, descubrir las relaciones entre los productos que los clientes tienden a comprar juntos o secuencialmente en el tiempo.
- Detección de desviaciones: Se trata de explicar por qué ciertos valores muestran desviaciones de la norma o de una expectativa previamente conocida.



### EJEMPLO PRÁCTICO

El mismo banco que anteriormente había decidido crear un data mart para su departamento de riesgos, se está planteando introducir capacidades de data mining. ¿Cómo podría beneficiarse de ellas?

### SOLUCIÓN

Probablemente en una situación de crisis, muchas empresas tengan la tentación de intentar defraudar al banco para conseguir préstamos que las puedan salvar.

Por este motivo, sería muy interesante aplicar el data mining a la gestión del fraude. De esta forma se podrían detectar anomalías en los datos aportados por los clientes mediante el análisis de los documentos que entregan como cuentas anuales, impuestos, etc. Por ejemplo, empresas que declaran una cantidad de ingresos anormalmente alta para lo que es su trayectoria histórica o su sector, etc.

De este modo, los analistas de riesgos podrían mejorar su toma de decisiones al saber que los datos que están analizando son fiables.

## 2.6 OLAP

El paradigma de aplicación decisional lo constituye el denominado proceso analítico on-line (OLAP) que consiste en la observación multidimensional de los hechos (compras, ventas, costes, gastos, ingresos,) que sean relevantes para una determinada área de negocio, departamento o ámbito de actividad.

El proceso analítico está condicionado a un cierto modelo de negocio multidimensional preestablecido para cada área o departamento. Este proceso debe ser totalmente interactivo, permitiendo una gran potencia de navegación por las estructuras, con entera libertad de movimientos sólo limitada por el propio ámbito del modelo y por los niveles de seguridad establecidos. Los destinatarios de estas aplicaciones son los analistas de información especializados, encargados de proporcionar datos de gran criticidad a los decisores del más alto nivel.

Para dar respuesta a estos requerimientos, el proceso debe contemplar el establecimiento de un modelo de negocio multidimensional para cada área o departamento, y su implementación a través de estructuras de datos optimizadas para el análisis. Estas estructuras pueden ser almacenadas y

gestionadas por gestores de bases de datos relacionales o multidimensionales, también denominados servidores OLAP.

El último eslabón de la cadena lo constituyen las herramientas de visualización OLAP, que disponen de una interfaz gráfica de usuario para permitir una potente navegación y explotación del modelo de una forma muy intuitiva. Estas herramientas permiten no sólo la visualización de información tabulada, sino también su representación gráfica en múltiples formatos en dos y tres dimensiones. La navegación OLAP se caracteriza por la posibilidad de analizar de forma totalmente interactiva cualquier porción del modelo multidimensional establecido. Entre las funciones más típicas se encuentran:

- Drill-up/drill-down, es decir, la posibilidad de descender a niveles jerárquicos inferiores de una estructura partiendo del elemento de un determinado nivel y explotando todos los elementos de los que se compone, y viceversa, la posibilidad de ascender a niveles más superiores, agregando los datos de los niveles inferiores.
- Drill-across, o la posibilidad de movimiento a través de los distintos elementos que componen un determinado nivel jerárquico de una estructura.
- Pivot, o la posibilidad de rotación de los ejes del análisis (dimensiones), para visualizar la información desde otra perspectiva.
- Slice & dice, es decir, la posibilidad de fijar el ámbito del análisis a una porción del modelo, bien por corte de una dimensión, estableciendo una "rebanada", bien por corte de múltiples dimensiones, estableciendo un cubo más pequeño.

Estas funciones están soportadas por el servidor OLAP. En el caso de no utilizar un servidor especializado, la herramienta de visualización OLAP se encargará de generar dinámicamente el cubo multidimensional en local, aunque el rendimiento de esta solución es bastante inferior porque debe descargarse una gran cantidad de información en local.

Existen productos en el mercado, denominados servidores ROLAP (OLAP relacional), que actúan como motores de acceso multidimensional a datos relacionales. Los rendimientos son mejores, porque los cubos se generan dinámicamente en el servidor y al cliente sólo viaja el resultado de la consulta.

El proceso OLAP sobre datos multidimensionales y relacionales se denomina HOLAP (OLAP híbrido). En algunas aplicaciones puede ser necesario descender a las operaciones de detalle para comprender el motivo por el cual se producen determinados comportamientos. En este caso, si disponemos de datos multidimensionales y relacionales, las herramientas deben ser capaces de acceder a ambos de forma integrada.

Finalmente, debemos mencionar que existen herramientas de usuario final de propósito general (las hojas de cálculo constituyen el ejemplo más característico) que permiten la explotación OLAP de la información. Algunos servidores OLAP incorporan una funcionalidad añadida que se instala en estas herramientas y que permiten la explotación OLAP de la información gestionada por estos servidores. Tienen la gran ventaja de aprovechar el conocimiento ofimático extendido entre los usuarios y no suponer una nueva inversión ni por adquisición ni por mantenimiento.

## 2.7 EIS

Los sistemas de información para la dirección (EIS), como su propio nombre indica, normalmente van dirigidos a la alta dirección, pero se aplican asimismo a los niveles ejecutivos intermedios y decisores de cualquier nivel con suficiente autonomía de negocio.

Estos sistemas se caracterizan por disponer de una interfaz de usuario perfectamente ajustada a la estrategia de negocio y al estilo de dirección de la compañía. El ideal de un EIS es ofrecer al gestor una sensación de absoluta libertad (dentro de los límites impuestos por el modelo de datos y el sistema de navegación) e inmediata interacción a través de funcionalidades para la consulta y la navegación. Esto se consigue si se dispone de toda la información ya calculada y su recuperación está totalmente optimizada. Todo ello implica un sistema de navegación preestablecido sobre unas vistas de información predefinidas, es decir, un diseño y construcción previos.

Los EIS son sistemas que pueden verse sometidos a frecuentes cambios debido a esta necesidad casi biológica de adecuación a las necesidades de la alta dirección. Esta es la razón por la que numerosos EIS se hayan visto abandonados con el tiempo. Para evitarlo, es imprescindible disponer de un buen modelo de datos que recoja toda la información susceptible de ser utilizada por el entorno de soporte a la decisión. Las sucesivas adaptaciones de los EIS sólo deben preocuparse de modificar la interfaz de presentación, sin necesidad de modificar el ciclo completo de obtención de la información y su transformación al modelo de negocio.

## 2.8 El sistema de reporting

Los almacenes de datos que se crean en la organización con el fin de facilitar la toma de decisiones constituyen una excelente plataforma para el reporting empresarial o departamental, según la amplitud del modelo

contemplado. A diferencia de las restantes modalidades de explotación, basadas en herramientas de acceso, el reporting ofrece un enfoque más gestionado y está basado en herramientas de suministro de información. Esto es debido a que el reporting siempre va asociado a un diseño y construcción previos y una estrategia de despliegue de los resultados.

Por lo tanto, esta solución presupone una planificación, diseño y construcción previos, adecuados a las necesidades corporativas o departamentales definidas.

Para su implementación, deben diseñarse dos estrategias, una primera de generación de información y otra de despliegue de resultados a los usuarios. La primera pone el énfasis en los contenidos y continentes (o fondos y formas), y la segunda en las personas y en los medios. Para el diseño e implementación de la estrategia de generación de información, se deben resolver las siguientes cuestiones:

- ¿Qué información debe ser generada?
- Obtención de la relación de informes que componen la solución.
- Descripción del objetivo de cada informe.
- Definición temporal de cada informe: imagen estática (actual o histórica) o evolución temporal.
- Descripción de la porción del modelo de datos afectado por cada informe.
- ¿Cuánta información debe ser generada?
- Definición de la frecuencia de generación de cada informe periódico.
- Descripción del grado de detalle y agregación requerido en cada informe.
- Definición del intervalo temporal de cada informe de evolución.
- ¿Cómo debe ser generada?
- Diseño de cada informe respetando una serie de estándares.
- Descripción del mecanismo de personalización de cada informe por usuario o grupo.
- Formato de cada informe.

Para el diseño e implementación de la estrategia de despliegue de resultados, se deben atender a los siguientes aspectos:

- ¿Quién accede a la información?
- Obtención de las relaciones grupo/usuario por informe.
- Descripción del mecanismo de publicación/suscripción.
- Descripción del mecanismo de lista de distribución.
- ¿Cuándo está disponible la información?
- Descripción del evento y mecanismo de disparo para cada informe.
- ¿Dónde está disponible la información?
- Descripción de directorios de resultados compartidos y personales.

- Descripción del mecanismo de mensajería.
- Descripción del mecanismo de publicación en web.
- Descripción del mecanismo de distribución a dispositivos especiales (teléfono móvil, tablet, etc.)

Dada la gran variedad de mecanismos y medios disponibles para la implementación de la solución, es posible dotar a la organización de un conjunto de herramientas de suministro de información, que, operando de forma coordinada, cumpla con los requisitos del sistema de reporting diseñado. En cualquier caso, la solución debe contemplar una herramienta que permita la construcción de informes (lenguaje de programación, herramienta de generación de informes, herramienta de desarrollo...).

## 2.9 Sistemas de apoyo a la decisión (DSS)

La competitividad de los mercados, las demandas de los consumidores, el desarrollo de la tecnología... son aspectos que hacen necesario tomar decisiones precisas y oportunas para garantizar el éxito de cualquier compañía. El proceso de toma de decisiones por lo general se basa en nuestro aprendizaje a través de la experiencia, nuestra intuición, creatividad...

Realmente, identificar y evaluar las posibles soluciones es un proceso que resulta complicado, que requiere el manejo de numerosas fuentes de información, el análisis de múltiples variables... y, todo ello, con la rapidez que nos exigen los vertiginosos cambios sociales. En la actualidad, el desarrollo de herramientas informáticas ha originado soluciones que facilitan la toma de decisiones como los sistemas de apoyo a la decisión, más conocidos como DSS.

Los DSS fueron definidos por Scott Morton en los años 70 de la siguiente manera:

*"Son sistemas interactivos basados en ordenadores, que ayudan a la toma de decisiones utilizando datos y modelos para resolver problemas no estructurados. [...] Los DSS asocian los recursos intelectuales de los individuos con las capacidades de los ordenadores para mejorar la calidad de las decisiones. Es un sistema de soporte basado en ordenadores para los responsables de las tomas de decisiones y que tratan con problemas semiestructurados."*

De forma genérica podemos decir que un DSS es cualquier sistema informático que facilita la toma de decisiones en una empresa. Sin

embargo, la expresión sistemas de apoyo a la decisión ha sido interpretada por muchos autores como una metodología conceptual. De forma que se puede considerar el DSS como un proceso y, al mismo tiempo, como una propia aplicación.

Se suelen utilizar sistemas de soporte independientes para marketing, finanzas y contabilidad; un sistema de gestión de la cadena de suministro (SCM), para producción, y diferentes sistemas expertos para producir diagnósticos para la reparación de productos. El DSS debe reunir a todos los sistemas anteriores.

Pero, ¿por qué utilizar DSS?, principalmente destacamos los siguientes motivos:

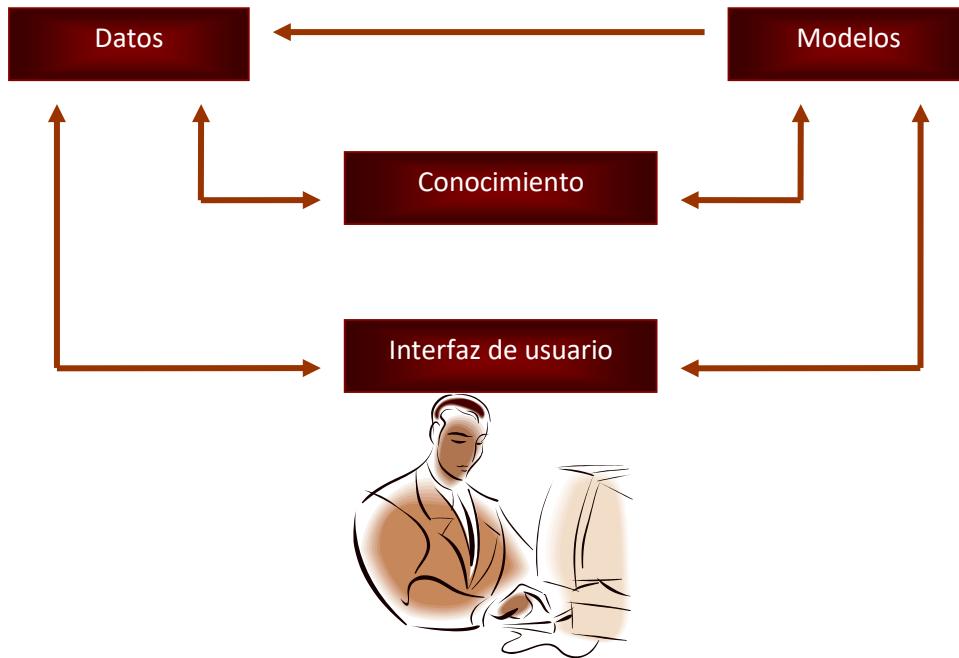
- Son rápidos.
- Se mejora la comunicación y colaboración.
- Contribuyen a elevar la productividad.
- Mejoran la gestión de datos.
- Sirven como soporte a la calidad.
- Superan los límites cognitivos en lo que se refiere al almacenamiento y gestión de los datos.
- Se pueden utilizar en cualquier lugar.

Una aplicación de DSS puede estar compuesta de los siguientes subsistemas:

- Gestión de datos.
- Gestión de modelos.
- Interfaz de usuario.
- Gestión basada en el conocimiento (SGC).

Los DSS necesitan contar con datos fidedignos para poder resolver problemas o para analizar oportunidades y estrategias de las que depende el éxito de una empresa. Estos datos pueden proceder de diferentes fuentes, incluyendo la web. Los datos son el primer componente de la arquitectura de un DSS.

Los datos relacionados con una situación específica se manipulan utilizando modelos de la realidad que debemos crear. Estos modelos pueden ser estándar o personalizados (customizados). Gracias al análisis de los datos podemos obtener el conocimiento que nos permita tomar las mejores decisiones. Por último, necesitamos desarrollar una interfaz que permita a los usuarios del sistema acceder a esos datos y modelos.

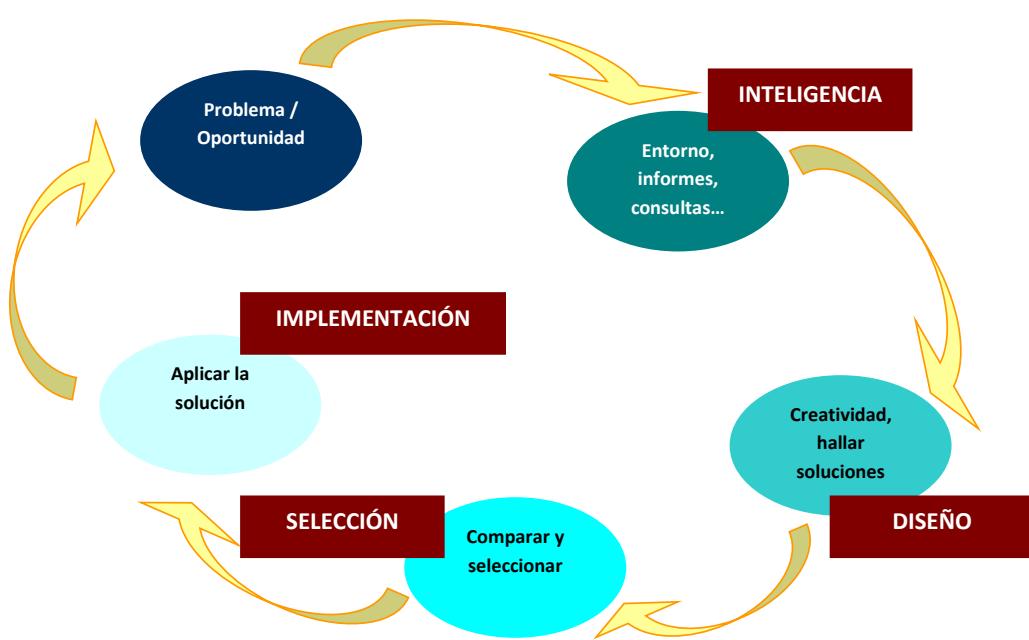


Modelo general de un sistema de soporte a la toma de decisión

Fuente: Elaboración propia

### **2.9.1 Fases de un proceso de toma de decisiones**

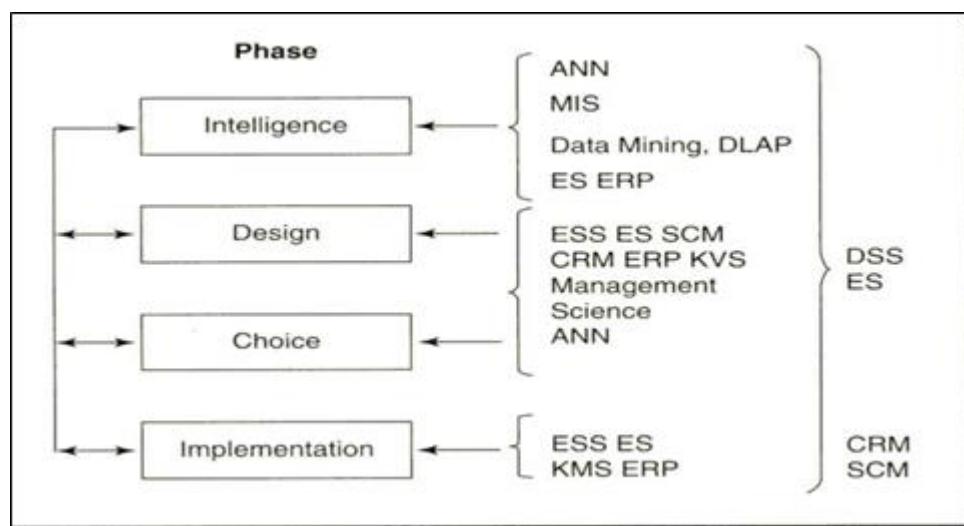
Las fases que podemos identificar dentro del proceso de toma de decisiones las podemos representar de la siguiente manera:



Fases del proceso de toma de decisión

Fuente: Elaboración propia

- **Inteligencia:** En esta fase el responsable de tomar las decisiones examina la realidad e identifica y define el problema. Se establece también la propiedad del problema.
- **Diseño:** Se construye un modelo que representa el sistema. Se realiza haciendo las hipótesis que simplifican la realidad y se establecen las relaciones entre las variables. El modelo se valida a continuación y se determinan los criterios para la evaluación de las diferentes alternativas identificadas. A veces, el proceso de desarrollo de modelos identifica soluciones alternativas y viceversa.
- **Selección:** Incluye la selección de una solución propuesta al modelo (no necesariamente al problema que se representa). La solución se comprueba para determinar su viabilidad. Cuando la solución propuesta parece razonable, se está listo para la última fase: implementación de la decisión (no necesariamente de un sistema). La implementación con éxito conduce a la solución del problema real. Los fallos conducen a un retorno a una fase anterior del proceso. De hecho, se puede volver a una fase anterior durante cualquiera de las tres últimas fases.
- **Implementación:** Implica poner en funcionamiento una solución recomendada, no necesariamente la implementación con un sistema informático.



Fases del proceso de toma de decisiones con las herramientas tecnológicas necesarias.

Fuente: Turban. Decision Support and Business Intelligence Systems. Editorial Pearson

## 2.9.2 Tipos de DSS

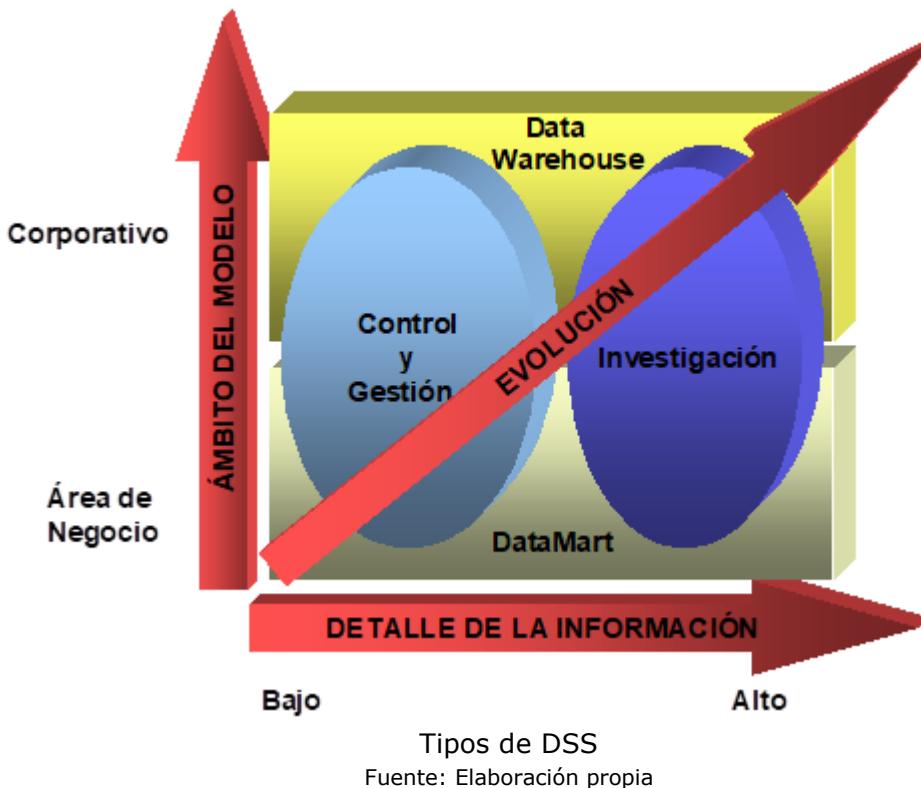
Si nos atenemos al detalle de la información que almacenan, nos encontramos con una forma de clasificar los sistemas decisionales (DSS):

- **Sistemas para el control y gestión.** La información de detalle no es relevante para el análisis. Se definen una serie de indicadores de rendimiento que se alimentan a los niveles requeridos, se planifica su comportamiento y posteriormente se realiza un seguimiento del comportamiento real de los mismos, observando las diferencias con los valores previstos, lo que proporciona una visión muy clara y concisa de la marcha del negocio en sus aspectos más críticos.
- **Sistemas para la investigación.** Son más genéricos, y su objetivo es el establecimiento de un almacén de datos lo suficientemente completo y detallado para permitir a especialistas en el negocio la exploración de los datos, el descubrimiento de patrones de comportamiento y la obtención de conclusiones que no resultan obvias.

Ambos sistemas pueden convivir en una misma instalación, cada uno de ellos destinado a cubrir unas determinadas necesidades. Podemos imaginarnos un almacén de datos donde se mantenga por un lado la información al máximo nivel requerido por los procesos de investigación, y por otro, esa misma información debidamente estructurada y agregada para soportar aplicaciones de control y gestión.

Estas estructuras agregadas se alimentarían de la información detallada. De hecho, la implementación de datamarts dependientes que se explicó anteriormente puede responder a estos mismos criterios.

La siguiente ilustración muestra las diferentes tipologías de sistemas de soporte a la decisión, atendiendo a la amplitud del modelo de negocio y al detalle de la información almacenada. También se observa la probable evolución que sufrirán dentro de la organización.



Podemos afirmar que los sistemas para el control y gestión están más orientados a la alta dirección, para que, de un solo vistazo a la información, pueden conocer las cifras más relevantes de la organización y actuar en consecuencia, alertando a los ejecutivos de cuentas, de producto, etc.

Los segundos están más orientados a quienes necesitan explorar los datos más detallados para analizar por qué se han producido determinados hechos e informar a la dirección. También serán utilizados por los especialistas del negocio para obtener conocimiento a partir de la información, en forma de patrones de comportamiento, predicción de evoluciones, etc.

Otra forma de clasificar los DSS es atendiendo a la actualidad de la información que almacenan y suministran. Los almacenes pueden contener información totalmente actualizada, o bien información "congelada" en el tiempo, con una frecuencia definida:

- Los primeros podríamos denominarlos sistemas de información tácticos, utilizados en las tomas de decisiones diarias. Así, un almacén de datos sobre clientes que recoja la situación actual de todas las cuentas pertenecientes a cada cliente permite decidir qué acciones de tipo comercial o de otra índole se deben emprender de forma inmediata sobre cada cliente en particular.

Parece claro que el director comercial de tu compañía necesita un sistema de soporte a la toma de decisiones que le ayude a priorizar su tiempo y, por tanto, decidir a qué leads dirige su oferta comercial

por ser potencialmente más rentables y estar más interesados y, por lo contrario, decidir qué leads es mejor abandonar.

- Los segundos están orientados al seguimiento de la actividad, mostrando evoluciones y tendencias que permiten a la organización la toma de decisiones estratégicas, como el abandono de cierta línea de productos, la potenciación de determinado canal de distribución, la ampliación de la estructura territorial...

Como puede apreciarse, no existe un único modelo de sistema decisional, sino que cada organización o departamento debe definir el suyo propio, de acuerdo con las necesidades planteadas.

### 3. BUSINESS ANALYTICS: CLUSTERING

*En el proyecto que estás desarrollando por encargo del director comercial de tu empresa, necesitas realizar un proceso de clustering para intentar descubrir perfiles de clientes dentro de la empresa. Desconoces si tu organización está considerando a todos los clientes por igual, al no haber realizado una segmentación previa, y piensas que mediante la búsqueda de clusteres podrás descubrir clientes que comparten determinadas características.*

*Una vez descubiertos esos clusteres, quieres determinar cuál o cuáles de ellos son más rentables para la empresa, para lo que buscas otros clientes en el mercado que cuenten con las mismas características y que, por tanto, son especialmente rentables para el negocio.*

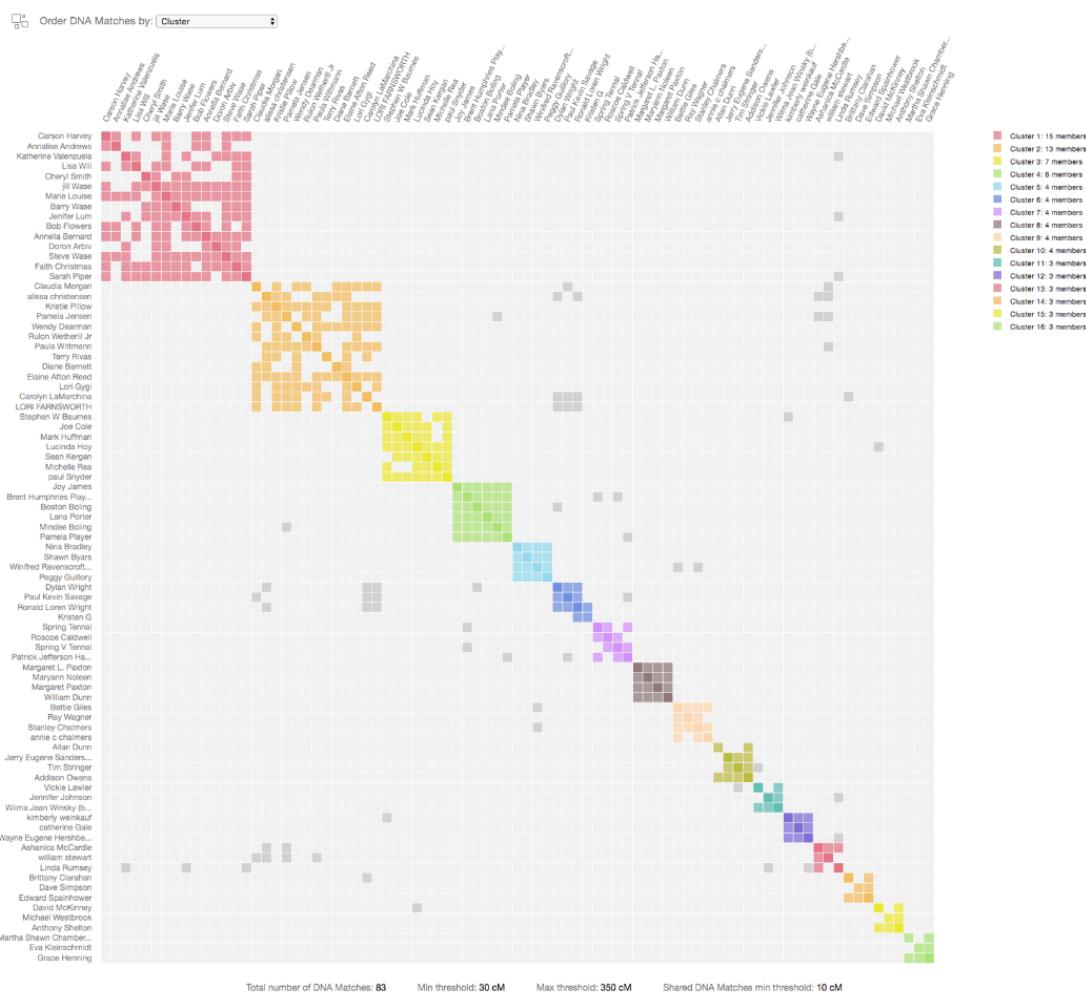
Clusterizar es un proceso realizado mediante la aplicación de un algoritmo a un conjunto de elementos, de los cuales se conocen diversas características, de modo que son agrupados automáticamente en dos o más grupos, denominados clusters, que comparten aspectos similares entre ellos y distintos a los elementos que forman parte de otro u otros cluster.

El uso del clustering se ha incrementado mucho en cuanto a su aplicación a todo tipo de ámbitos, por la facilidad de uso que tiene el machine learning en la actualidad. En el caso del clustering, el aprendizaje automático aplicado es el denominado de tipo no supervisado, ya que, partiendo de un conjunto de datos de entrada, pretendemos obtener como salida información sobre su estructura que nos resulta desconocida, pero sin definir a priori ninguna clase predeterminada como ocurre cuando hablamos de clasificación en vez de clusterización.

Clustering es diferente a clasificación. La diferencia es que la clasificación corresponde al aprendizaje supervisado, a diferencia del clustering que ya

se ha mencionado que es aprendizaje no supervisado. Esto quiere decir que en los problemas de clasificación disponemos de información sobre las variables de entrada y de salida. El algoritmo de clasificación más sencillo sería el de clasificación binaria, por ejemplo, determinar si una imagen representa a una persona o no.

Un ejemplo de clustering complejo sería agrupar a un conjunto de individuos, en base a los datos de su ADN, por el grado de familiaridad que presentan.



Clustering de un grupo de personas por su vinculación de familiaridad  
Fuente: Myheritage.com

Clusterizar es de gran ayuda cuando, sin tener un conocimiento previo, queremos saber cómo están organizados un conjunto de datos.

Después de haber aterrizado en apartados previos el concepto de business intelligence y de haber visto una introducción a los modelos típicos de arquitectura para establecer una plataforma de BI, es momento de entrar a

explorar dos herramientas útiles y de alto valor para cualquier aplicación inicial de BI.

Estos métodos se conocen en inglés como clustering y classification (agrupación y clasificación). En este apartado veremos sus usos, métodos de aplicación y forma de utilización.

Básicamente la agrupación es la tarea de dividir la población o los puntos de datos en varios grupos, de forma que los puntos de datos en un grupo sean parecidos a otros puntos de datos en el mismo grupo. En otras palabras, el objetivo es separar grupos con características similares y ordenarlos en grupos.

Comprendamos esto con un ejemplo típico de la analítica empresarial. Supongamos que eres el responsable de una tienda de alquiler y quieres comprender las preferencias de tus clientes para ampliar el negocio.

¿Es posible que estudies los detalles de cada cliente y diseñas una estrategia comercial única para cada uno de ellos?... En principio, salvo que tengas un negocio muy pequeño y tus clientes se cuenten con los dedos de las manos, parece que no es posible.

Pero, lo que, si puedes hacer, es agrupar a todos tus clientes, aunque sean miles o cientos de miles, en diez grupos según sus tendencias de consumo y utilizar una estrategia separada para los clientes en cada uno de estos los grupos.

### 3.1 Tipos de clustering

*Tienes una idea para sorprender a tu director comercial. Crees que seguro que el resultado de la clusterización de vuestros clientes históricos le parece muy interesante. Eso sí... para interpretar el conocimiento derivado de los clusteres hay que tener conocimientos de negocio. Por este motivo le pides al director comercial que te deje contar con el conocimiento de alguno de sus analistas o técnicos de marketing y de ese modo poder interpretar adecuadamente los resultados expresados por la clusterización.*

Una vez conocido y comprendido el esquema general o la visión de lo que se realiza en el clustering, vamos a revisar los tipos principales de clustering:

- Hard clustering / Agrupación dura: En este tipo de agrupación, cada punto de datos puede pertenecer o no completamente a un clúster.
- Soft Clustering/ Agrupación suave: En la agrupación suave, por el contrario, no se asigna un punto de datos de forma taxativa a un

grupo, sino que se busca asignar una probabilidad de que el dato pertenezca a uno u otro grupo (clúster).

### 3.2 Tipos de algoritmos de clustering

Dado que la tarea de agrupamiento es subjetiva, los medios que se pueden utilizar para lograr este objetivo son muchos. Cada metodología sigue un conjunto diferente de reglas para definir la similitud entre los puntos de datos. De hecho, hay más de cien algoritmos de agrupación conocidos. Pero son muchos menos los algoritmos que se usan de forma generalizada. Veámoslos en detalle:

- Modelos de conectividad: Como su nombre indica, estos modelos se basan en la noción de que los puntos de datos más cercanos en el espacio de datos exhiben más similitud entre sí que los puntos de datos que se encuentran más lejos.

Estos modelos pueden seguir dos enfoques:

- En el primer enfoque, comienza con la clasificación de todos los puntos de datos en grupos separados y luego los agregan a medida que disminuye la distancia.
- En el segundo enfoque, todos los puntos de datos se clasifican como un solo grupo y luego se dividen a medida que aumenta la distancia. Además, la elección de la función de distancia es subjetiva.

Estos modelos son muy fáciles de interpretar, pero carecen de escalabilidad para manejar grandes conjuntos de datos. Ejemplos de estos modelos son el algoritmo de agrupamiento jerárquico y sus variantes.

- Modelos centroides: Estos algoritmos de agrupación iterativos son aquellos en los que la noción de similitud se deriva de la cercanía de un punto de datos al centroide de los grupos. El algoritmo de agrupación de K-Means es un algoritmo popular que se incluye en esta categoría.

En estos modelos, el número de grupos (clústeres) deseados a la finalización del algoritmo, deben mencionarse de antemano, lo que hace que sea importante tener un conocimiento previo del conjunto de datos (o en su defecto, clusterizar con un número de agrupaciones diferentes y comparar los resultados para decidir cuál aporta más conocimiento sobre el negocio).

Estos modelos se ejecutan iterativamente para encontrar los óptimos locales.

- Modelos distribuidos: Estos modelos de agrupamiento se basan en la noción de cuánto probable es que todos los puntos de datos en el grupo pertenezcan a la misma distribución (por ejemplo, la distribución normal o gaussiana, etc.).

Estos modelos a menudo sufren de sobreajuste. Un ejemplo popular de estos modelos es el algoritmo de maximización de expectativas, que utiliza distribuciones normales multivariadas.

- Modelos de densidad: Estos modelos buscan en el espacio de datos áreas de densidad variada de dichos puntos de datos. El algoritmo aísla varias regiones de densidad diferentes y asigna los puntos de datos dentro de estas regiones en el mismo grupo. Ejemplos populares de modelos de densidad son DBSCAN y OPTICS.

### **3.2.1 Método K Clustering**

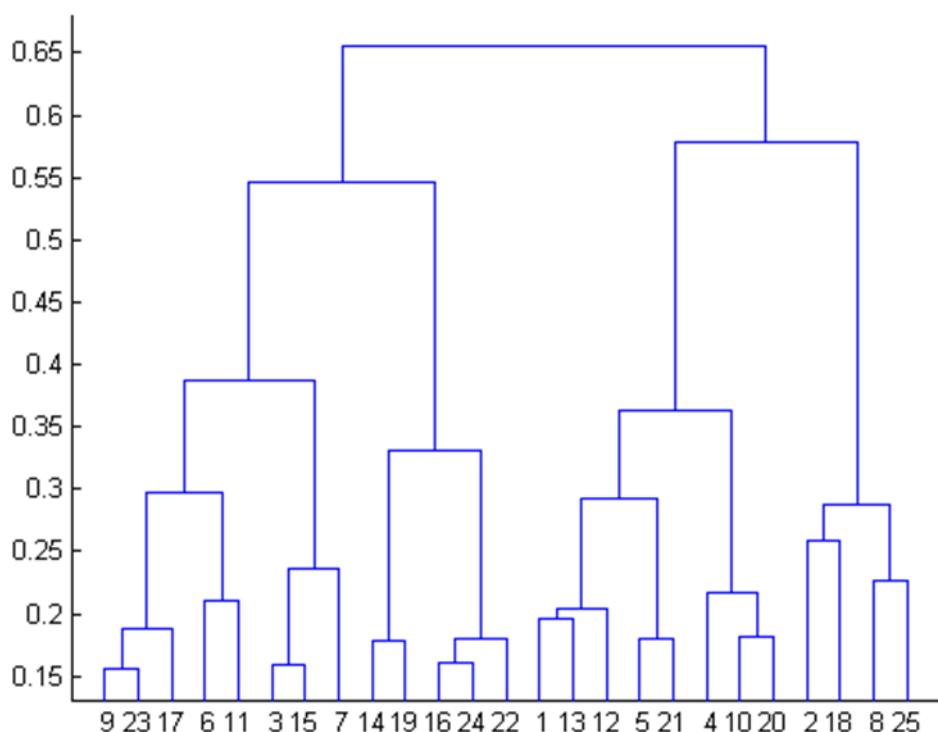
K means, o K medias, es un algoritmo de agrupamiento iterativo, que tiene como objetivo encontrar máximos locales en cada iteración. Este algoritmo funciona en estos cinco pasos:

- Se especifica el número K de clústeres a crear.
- Aleatoriamente, se asigna cada punto de datos a un clúster.
- Se calculan los centroides del clúster
- Se repite iterativamente el cálculo de los centroides de los cluster.

### **3.2.2 Método jerárquico de agrupación / Hierarchical clustering**

La agrupación jerárquica, como su nombre indica, es un algoritmo que construye una jerarquía de agrupaciones. Este algoritmo comienza con todos los puntos de datos asignados a un grupo propio. Luego, dos grupos más cercanos se fusionan en el mismo grupo. Al final, este algoritmo termina cuando solo queda un clúster.

Los resultados de la agrupación jerárquica se pueden mostrar utilizando el dendrograma.



Dendrograma en el método jerárquico de agrupación

Fuente: Elaboración propia

### **3.2.3 Diferencias entre el método de clustering K y el jerárquico**

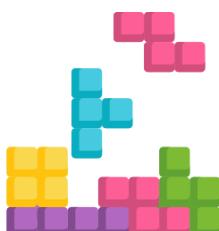
Tras haber visto una introducción de cómo funcionan los algoritmos para la agrupación de datos, veamos una pequeña lista de las principales diferencias entre estos dos populares métodos:

- La agrupación jerárquica no puede manejar bien los grandes datos, pero la agrupación K sí.
- En la agrupación K, dado que comenzamos con la elección aleatoria de agrupaciones, los resultados producidos al ejecutar el algoritmo varias veces pueden diferir. Mientras que los resultados son reproducibles en la agrupación jerárquica.
- Se da la circunstancia que el K clustering funciona bien cuando la forma de los grupos es hiperesférica (como un círculo en 2D, una esfera en 3D).
- La agrupación K requiere conocimiento previo de K, es decir, del número de grupos en los que se desea dividir el conjunto de elementos. Mientras que el método jerárquico puede detenerse en cualquier cantidad de grupos que se considere apropiados interpretando el dendrograma de forma adecuada.

### 3.2.4 Aplicaciones comunes de los algoritmos de agrupación

La agrupación o clusterización tiene un gran número de aplicaciones a lo largo y ancho de varios dominios. Algunas de las aplicaciones más populares de clustering son:

- Motores de recomendación.
- Segmentación de mercado.
- Análisis de redes sociales.
- Agrupación de resultados de búsqueda.
- Imágenes médicas.
- Segmentación de imagen.
- Detección de anomalías.



#### EJEMPLO PRÁCTICO

¿Cómo podría un banco aprovechar los algoritmos de clusterización para mejorar la comercialización y marketing de sus productos?

#### SOLUCIÓN

El banco podría tomar todos los datos de los que dispone de sus clientes, tanto aquellos sobre cada usuario (edad, población, etc.) como los datos sobre su comportamiento bancario (ingresos recibidos, transferencias realizadas, etc.)

Aplicando un algoritmo de clustering, le permitiría encontrar agrupaciones con características comunes que probablemente no ha tenido en cuenta a la hora de realizar su marketing. Es decir, es muy probable que los segmentos de clientes que había definido el banco a priori sean muy diferentes a los grupos que devuelve en análisis cluster, con lo que deberían redefinir muchas políticas comerciales.

### 3.2.5 Ejemplo: Clusterizar ingresos de hogares mediante K clusteres

Vamos a partir de un conjunto de familias que tienen unos ingresos brutos anuales determinados. Tomemos hogares con ingresos anuales de 40.000 €, 55.000 €, 70.000 €, 100.000 €, 115.000 €, 130.000 € y 135.000€.

Y vamos a agrupar los hogares en dos grupos. Tomando sus ingresos como medida de similitud, entonces el primer grupo tendría como miembros a los hogares que ganan 40.000 €, 55.000 €, 70.000€ y el segundo grupo tendría los hogares que ganan 100.000 €, 115.000 €, 130.000 €, 135.000 €.

El procedimiento que nos permite llegar a la agrupación que se ha indicado en dos clusteres se debe a que 40.000 y 135.000 son los más alejados entre sí, y requerimos tener dos cúmulos, así que tienen que estar en los diferentes cúmulos.

55.000 está más cerca de 40.000 que de 135.000, así que 40.000 y 55.000 estarán en el mismo grupo. Del mismo modo, 130.000 y 135.000 estarán en el mismo grupo.

70.000 está más cerca de 40.000 y 55.000 que de 130.000 y 135.000, por lo que 70.000 debería estar en el grupo con 40.000 y 55.000.

115.000 está más cerca de 130.000 y 135.000 que del primer cúmulo con 40.000, 55.000 y 70.000, por lo que estará en el segundo grupo.

Finalmente, 100.000 está más cerca del segundo cúmulo con 115.000, 130.000 y 135.000, así que estará allí.

Por lo tanto, finalmente el primer cúmulo contendrá los hogares con ingresos de 40.000 €, 55.000 € y 70.000 €; y el segundo cúmulo contendrá los hogares con ingresos de 100.000 €, 115.000 €, 130.000 € y 135.000 €.

Podemos decir que, dado que el clustering agrupa elementos con propiedades similares, entonces la asignación de un elemento a un cluster puede ser vista como una forma de clasificación.

Correspondería a un científico de datos interpretar el resultado de la agrupación y los qué la clasificación puede significar. En este caso de ejemplo, la interpretación es realmente trivial:

- Aquí el grupo de los hogares con los ingresos anuales de 40.000 €, 55.000 € y 70.000 € representan una clase de hogares con bajos ingresos.
- El segundo grupo, conteniendo a los hogares con los ingresos anuales de 100.000 €, 115.000 €, 130.000 € y 135.000 €, representa una clase de los hogares con un alto ingreso.

Esta aproximación a como agrupa el algoritmo k-means los hogares en dos grupos se ha realizado de manera informal, basándonos en la intuición y el sentido común, pero dado que hay algoritmos de clusterización que agrupan los datos según las reglas precisas, veamos cómo funcionan.

Entre estos algoritmos, aparte de k-means se encuentran los algoritmos fuzzy-C-means, el algoritmo de agrupación jerárquica, el algoritmo de clustering gaussiano, la agrupación de umbral de calidad y el algoritmo de

agrupación de K-means, que ya conocemos cómo se aplica intuitivamente y ahora vamos a ver su basamento matemático del siguiente modo.

El algoritmo de agrupación k-means clasifica los puntos dados en k grupos, de tal manera que la distancia entre los miembros del mismo grupo se reduce al mínimo.

El algoritmo de agrupación de k-means determina los k-centroides iniciales (puntos que deben estar en el centro del cluster), asignando uno para cada grupo.

Posteriormente, cada elemento se clasifica en el cluster cuyo centroide es lo más cercano a ese elemento. Después de clasificar todos los elementos, se habrán formado los primeros k clusters.

Para cada cluster, recalculamos el centroide para que sea la media de los puntos de ese cluster. Después de mover los centroides, volvemos a calcular las clases. De tal forma que los elementos pueden cambiar de clase.

Entonces, tendremos que volver a calcular los centroides. Si los centroides no se mueven de nuevo, entonces significa que el algoritmo de agrupación de k-means ha terminado su trabajo.

Escoger los k-centroides iniciales es sencillo. Tan sencillo como que podríamos escoger los k-centroides al azar. Pero idealmente, nos vendría bien escoger puntos que pertenezcan a clusteres diferentes desde el principio.

Por lo tanto, será una buena idea el intentar maximizar su distancia mutua de algún modo. Simplificando el proceso, podríamos elegir el primer centroide como cualquier punto de los elementos. El segundo podría ser el que esté más lejos del primero. El tercero podría ser el que esté más lejos, tanto del primero, como del segundo, y así sucesivamente.

Calcular el centroide de un cluster dado es sencillo, puesto que un centroide de un cluster es sólo un promedio de los puntos de un cluster. Si un cluster contiene puntos unidimensionales con las coordenadas  $x_1, x_2, \dots, x_n$ , entonces el centroide de ese cluster será  $(1/n)*(x_1+x_2+\dots+x_n)$ .

Si un cluster contiene puntos bidimensionales con las coordenadas  $(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$ , entonces, la coordenada x del centroide del cluster tendría valor  $(1/n)*(x_1+x_2+\dots+x_n)$ ; la coordenada y tendría el valor  $(1/n)*(y_1+y_2+\dots+y_n)$ .

Este cálculo se generaliza fácilmente a dimensiones superiores. Si el valor de los elementos de mayores dimensiones en su coordenada  $x$  son  $x_1, x_2, \dots, x_n$ , entonces el valor en la coordenada  $x$  para el centroide es  $(1/n)*(x_1+x_2+\dots+x_n)$ .

Vamos a aplicar matemáticamente este algoritmo para clusterizar los hogares de nuestro ejemplo según sus ingresos familiares. Para aplicar el algoritmo de agrupación  $k$  en el ejemplo de los ingresos del hogar recordemos que tenemos hogares con ingresos de 40.000, 55.000, 70.000, 100.000, 115.000, 130.000 y 135.000 euros.

El primer centroide que se elige puede ser cualquier elemento, por ejemplo, 70.000. El segundo centroide debería ser el elemento que está más lejos de la primera elección, que es 135.000 ya que  $135.000 - 70.000 = 65.000$ , que es la mayor diferencia entre cualquier otro elemento y 70.000.

Por lo tanto, 70.000 es el centroide del primer cluster. 135.000 es el centroide del segundo cluster.

Ahora, 40.000, 55.000, 70.000 y 100.000 están más cerca de 70.000, tomando la diferencia, que, de 135.000, así que estarán en el primer cluster. Los elementos 115.000, 130.000 y 135.000 están más cerca de 135.000 que de 70.000, así que estarán en el segundo cluster.

Después de clasificar los elementos según los centroides iniciales, volvemos a calcular los centroides.

El centroide del primer cluster es  $(1/4)*(40.000+55.000+70.000+100.000) = (1/4)*265.000=66.250$ .

El centroide del segundo cúmulo es  $(1/3)*(115.000+130.000+135.000)=(1/3)*380.000 \sim 126.660$ .

Usando los nuevos centroides, reclasificamos los elementos como sigue:

- El primer cúmulo con el centroide 66.250 contendrá los elementos 40.000, 55.000, 70.000.
- El segundo grupo con el centroide 126.660 contendrá los elementos 100.000, 115.000, 130.000, 135.000.

Fijémonos que el elemento 100.000 se movió del primer cluster al segundo, ya que ahora está más cerca del centroide del segundo cluster (distancia  $|100.000-126.660|=26.666$ ) que del centroide del primer cluster (distancia  $|100.000-66.250|=33.750$ ). Como los miembros de los cluster cambiaron, tenemos que volver a calcular los centroides:

- El centroide del primer cúmulo es  $(1/3)*(40.000+55.000+70.000)=(1/3)*165.000=55.000$ .

- El centroide del segundo cúmulo es  
$$(1/4)*(100.000+115.000+130.000+135.000) =$$
$$(1/4)*480.000=120.000.$$

Usando estos centros, reclasificamos los elementos en los cúmulos:

- El primer centroide, 55.000 contienen los elementos 40.000, 55.000 y 70.000.
- El segundo centroide, 120.000 contendrá los elementos 100.000, 115.000, 130.000 y 135.000. Así, al actualizar los centroides, los cúmulos no cambiaron.

Por tanto, dado que sus centroides seguirán siendo los mismos, concluimos que el algoritmo termina con los dos cluster:

- El primer cluster, que tiene los elementos 40.000, 55.000, 70.000
- El segundo grupo que contiene los elementos 100.000, 115.000, 130.000 y 135.000.

### **3.2.6 Clustering con Power BI**

Cuando en Power BI utilizamos gráficos de dispersión, una de las posibilidades disponibles es la clusterización. El número de clusters que deseamos obtener es configurable. Según el volumen del data set, el tiempo de procesamiento del algoritmo será más o menos dilatado.



**VIDEOTUTORIAL**

En este vídeo puedes ver cómo realizar un cluster con Power BI:

<https://vimeo.com/user64513894/review/447098256/f3e39>

[1732e](#)

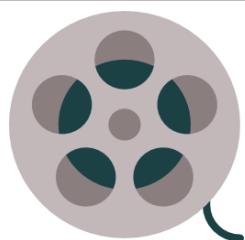
## 4. BUSINESS ANALYTICS: CLASIFICACIÓN

Analizando el proyecto en el cual estás involucrado, un planteamiento muy interesante puede ser el de clasificar en grupos homogéneos a los clientes históricos de la tienda de muebles. De esta forma, si consigues definir un segmento de clientes de alta rentabilidad, podrás conocer qué atributos en común tienen.

Estos atributos podrían ser por ubicación según código postal, edad, profesión, etc. De esta forma, cuando conozcas el perfil de los clientes más interesantes podrás compararlo con la base de datos de leads que habéis adquirido y, de ese modo, priorizar la campaña comercial sobre los clientes potenciales más similares a los clientes que ya sabéis que son rentables.

Una forma de realizar esto es utilizando la herramienta de business intelligence Power BI de Microsoft.

Dado que ya hemos estudiado varios algoritmos de clasificación en otras unidades, vamos en este momento a conocer la aplicación de un modelo de clasificación a una problemática empresarial muy común que es la clasificación de clientes ABC.



### VIDEO DE INTERÉS

En este vídeo podrás ver el cálculo dinámico de ventas (ABC de clientes)

<https://www.youtube.com/watch?v=NSuI8apdj3Y>

### 4.1 Introducción a Power BI

Actualmente, es necesario prestar mucha atención al tratamiento adecuado de los datos en el análisis de negocio para interpretar la información de la manera más precisa posible. Power BI es una herramienta creada por Microsoft que facilita el tratamiento y visualización de la información.



Power BI  
Fuente: Microsoft.com

Power BI permite la creación de consultas, modelos tabulares e informes con los que podemos interactuar y que, por lo tanto, resultan mucho más dinámicos que los informes realizados con otras herramientas.

El origen de Power BI se encuentra en los servicios de inteligencia de negocio integrados en SQL Server 2005. En 2010, Microsoft Excel incorpora componentes adicionales de business intelligence (Power Pivot), que propiciaron la aparición y desarrollo de Power BI. En la siguiente edición de Excel, 2013, ya se incluyó Power Query (herramienta de transformación de datos), Power View (para la creación de visualizaciones en Excel) y Power Map (para la visualización de datos sobre mapas).

Power BI fue lanzado por primera vez al público en general el 24 de julio de 2015 para convertir orígenes de datos diferentes y no relacionados en un modelo coherente, listo para visualizar y descubrir detalles analíticos.



#### ENLACE DE INTERÉS

En el siguiente enlace puedes profundizar un poco más en la historia de Power BI

<https://www.powerpro.consulting/post/2017-04-21-breve-historia-de-microsoft-power-bi>

## 4.2 Componentes de Power BI Los componentes de Power BI son:

- Power BI Desktop, que es una aplicación gratuita de escritorio Windows con la que podemos modelar los datos con facilidad y analizar los datos para encontrar patrones.
- Power BI pro, que funciona como "software as a service" en línea, de forma que es posible compartir los informes con otras personas del mismo equipo y trabajar de forma colaborativa cuando creamos nuevos informes y paneles.



### ENLACE DE INTERÉS

En el siguiente enlace se explican las diferencias entre Power BI Desktop y Power BI Pro

<https://www.innovaconsulting.es/blog/power-bi-o-power-bi-pro/>

- Power BI Mobile es una aplicación para móviles Windows, iOS y Android que nos permite visualizar los informes creados con Power BI Desktop o Power BI Pro, es decir, nos permite consumir los informes y cuadros de mando creados por los dos anteriores.

De manera sintética podemos decir que Power BI está formado por:

- Un conjunto de datos utilizados para mostrar las visualizaciones. Los conjuntos pueden ser sencillos o combinados. Power BI cuenta con muchos conectores, lo que nos permite tener acceso a gran cantidad de orígenes de datos.
- Las visualizaciones que se crean a partir de los conjuntos de datos. Las visualizaciones pueden ser sencillas, como una tarjeta que muestra un único valor, o más complejas, como gráficos, tablas y matrices que incluyen funcionalidades analíticas. A través de las visualizaciones presentaremos los datos de la manera adecuada.
- Los informes son un conjunto de visualizaciones y elementos relacionados y organizados en páginas. Se pueden crear desde Power BI Desktop, con un modelo de datos integrado en la aplicación, y desde Power BI Pro, que consume modelos de datos ya publicados en la nube. Los informes se pueden compartir desde el servidor de informes.
- Los paneles se crean desde Power BI Pro y son conjuntos de elementos de uno o más informes que se visualizan juntos, por lo

que permiten ofrecer mejor idea de la historia que estamos intentando representar.

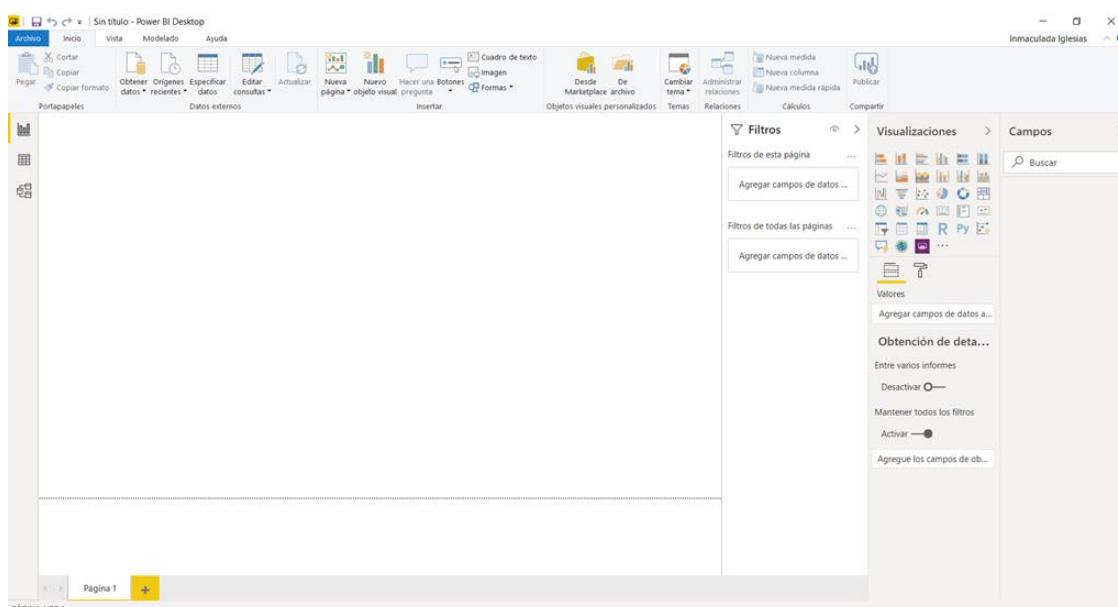
En función del trabajo que tengamos que realizar, en Power BI podemos diferenciar varios roles diferentes:

- El diseñador de conjuntos de datos es el encargado de la conexión, extracción, transformación, carga y modelado de datos.
- El diseñador de informes es el responsable de dar vida a los informes utilizando tanto las visualizaciones como el modelo de datos.
- El administrador Power BI supervisa las tareas y controla la gestión de recursos.
- Cuando utilizamos Power BI Pro encontramos el rol de gestor de la colaboración, que se encarga de tareas de configuración del almacenamiento, tareas de colaboración y seguridad en el acceso a los elementos almacenados en la nube.

### 4.3 Primeros pasos con Power BI Desktop

Power BI Desktop es una aplicación gratuita que podemos descargar desde su página oficial de manera segura.

En el entorno de trabajo de Power BI Desktop nos encontramos una distribución de los elementos similar a la de otras aplicaciones de Microsoft.

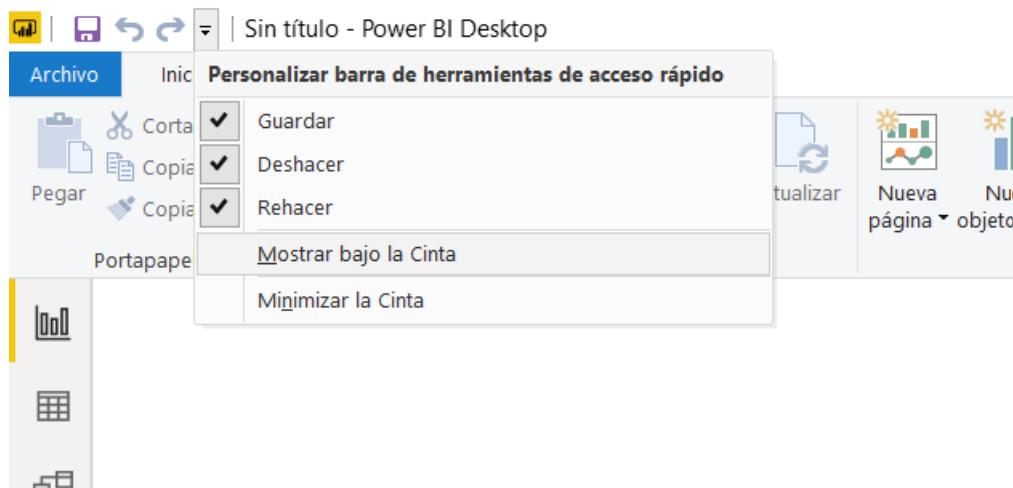


Vista principal de Power BI

Fuente: Elaboración propia

En la parte superior de la pantalla nos aparece la barra de herramientas de acceso rápido, en la que podemos seleccionar que opciones queremos que

aparezcan. Por defecto, tendremos seleccionadas las opciones guardar, deshacer yrehacer.



Barra de herramientas de acceso rápido

Fuente: Elaboración propia



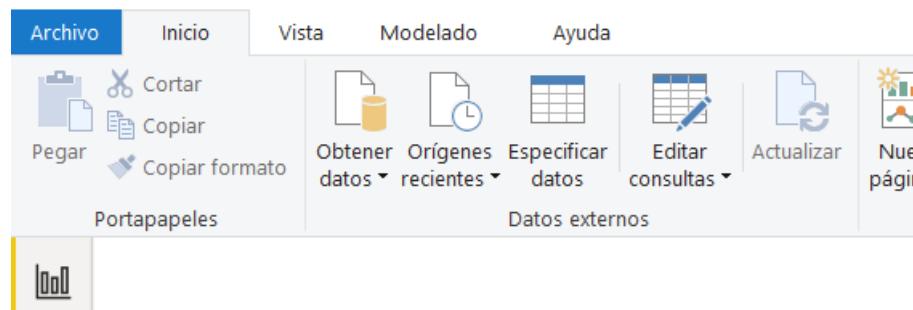
### VIDEOTUTORIAL

En este vídeo te explicamos como descargar e instalar Power BI Desktop:

<https://vimeo.com/user64513894/review/447099206/bd87>

[22c193](#)

Bajo la barra de herramientas de acceso rápido nos encontramos la cinta de opciones. En Power BI esta cinta está formada por cinco opciones: Archivo, inicio, vista, modelado y ayuda.

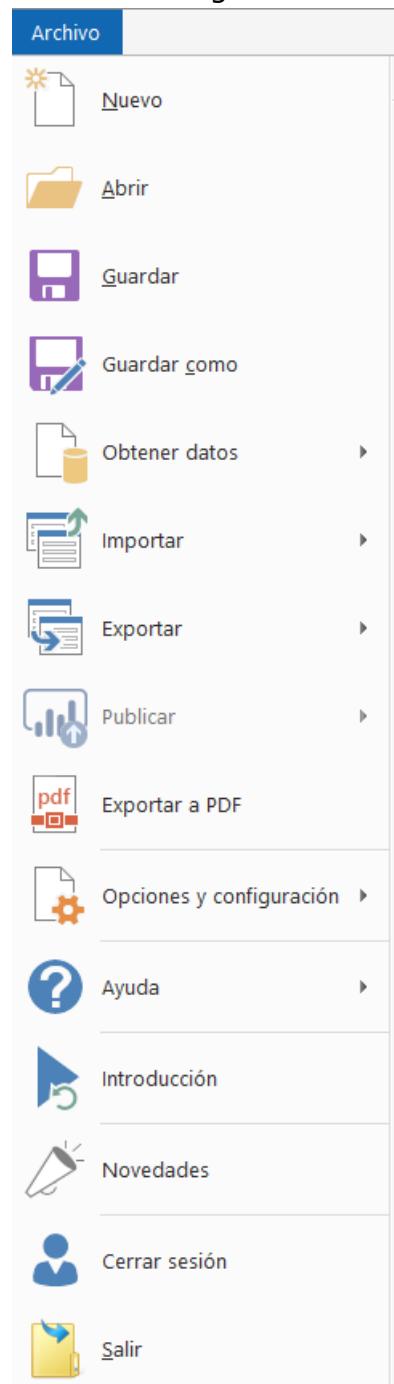


Cinta de opciones

Fuente: Elaboración propia

Las opciones de esta cinta se organizan en grupos, de forma que nos encontramos los siguientes grupos en cada una de las opciones:

- **Archivo.** Esta opción despliega un menú a través del cual accedemos a las opciones relacionadas con abrir o crear archivos de Power BI, obtener los datos de fuentes externas, importar o exportar la información y las opciones de configuración.

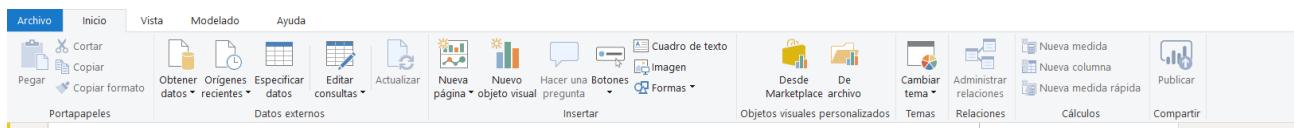


Menú archivo en Power BI

Fuente: Elaboración propia

- **Inicio.** Se divide en varios grupos de opciones:
  - Portapapeles, que tiene una finalidad similar a la que tiene en las aplicaciones de Office.

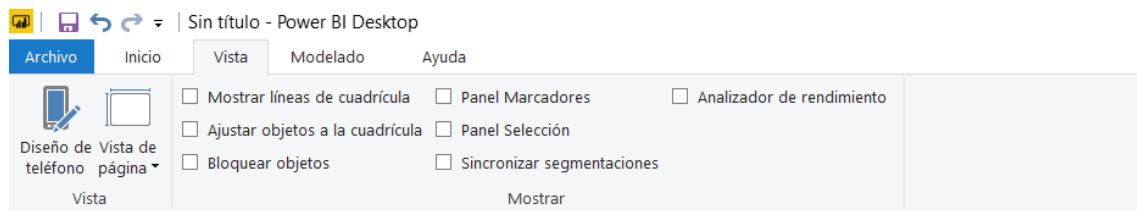
- Datos externos, un grupo fundamental, ya que nos permitirá seleccionar las fuentes de datos y acceder a Power Query donde podremos modelar los datos.
- Insertar, que nos permitirá añadir elementos al informe que estamos creando.
- Objetos visuales personalizados, que nos permite acceder al marketplace de Power BI para añadir nuevas visualizaciones.
- Temas, con los que podemos aplicar conjuntos de colores y tipografías predefinidos a nuestros informes.
- Relaciones, donde podremos ver las relaciones entre las tablas de nuestro modelo de datos.
- Cálculos, que usaremos para crear nuevas columnas a partir de los datos que tenemos almacenados.
- Compartir, para publicar nuestros informes.



Cinta de opciones de Inicio en Power BI

Fuente: Elaboración propia

- **Vista.** Nos permite definir cómo queremos visualizar el informe que estamos creando y nos da la opción de añadir paneles de opciones adicionales como como el panel de marcadores o el panel de selección.



Menú archivo en Power BI

Fuente: Elaboración propia

- **Modelado.** Este conjunto de opciones resulta fundamental porque nos permite crear nuevos datos, columnas o tablas a partir de los datos que estamos manejando. Encontramos los siguientes grupos de opciones:
  - Relaciones. Desde esta opción podemos visualizar y administrar las relaciones entre las tablas del modelo de datos.
  - Cálculos. Nos permite crear medidas, columnas y tablas a partir de nuestros datos.

- Hipótesis. Podemos crear un parámetro de hipótesis cuyo valor se puede establecer en la segmentación.
- Ordenar. Nos facilita la ordenación de columnas.
- Formato. Para modificar el tipo y formato de los datos del modelo.
- Propiedades. Establecemos cómo queremos resumir una columna de forma predeterminada o clasificamos los tipos de datos.
- Seguridad. Esta opción nos permite determinar los roles de acceso al informe para garantizar que cada persona que lo visualiza tiene acceso a la información precisa.
- Grupos. Podemos crear grupos para combinar varios valores en uno.
- Calendarios. Opción disponible cuando tenemos tablas de fechas.
- Preguntas y respuestas. Nos permite revisar las preguntas que formulan los usuarios en relación a los informes y administrar los términos y definiciones para ayudar a que los informes se entiendan mejor.



Cinta de opciones de Modelado en Power BI

Fuente: Elaboración propia

- **Ayuda.** Como en todas las aplicaciones de Microsoft, contamos con un grupo de opciones que nos ayudará a familiarizarnos con la herramienta que incluye ejemplos y enlaces a distintas páginas de la comunidad de Power BI.



Cinta de opciones de Ayuda en Power BI

Fuente: Elaboración propia

Además, en Power BI contamos con paneles de opciones en los laterales del espacio de trabajo.

A la izquierda de la pantalla tenemos tres opciones que nos permiten definir qué es lo que queremos visualizar: el informe, los datos o las relaciones.

A la derecha de la pantalla tenemos tres paneles de opciones:

- Filtros, donde definimos los filtros que aplicamos a cada página o elemento del informe.
- Visualizaciones, donde seleccionamos de qué forma queremos visualizar los datos.
- Campos, donde seleccionamos qué campos queremos visualizar del modelo de datos.

En la parte inferior de la pantalla seleccionamos la página del informe en la que vamos a trabajar y podemos añadir nuevas páginas.



### VIDEOTUTORIAL

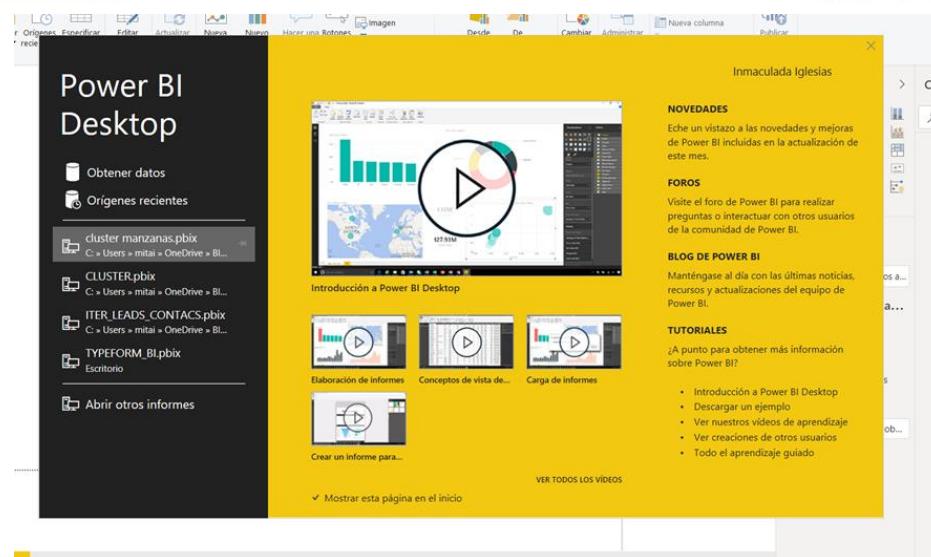
Con este vídeo entenderás mejor cómo se distribuyen las opciones en el espacio de trabajo de Power BI.

<https://vimeo.com/user64513894/review/447100043/531fb418a4>

## 4.4 Conectar con archivos de datos y transformar los datos

El entorno de trabajo de Power BI que nos permitirá conectar con los datos y modelarlos se llama Power Query. Power Query nos va a servir para conectar a los orígenes de datos, crear las consultas y transformar la estructura importada a las necesidades del informe que vamos a crear posteriormente.

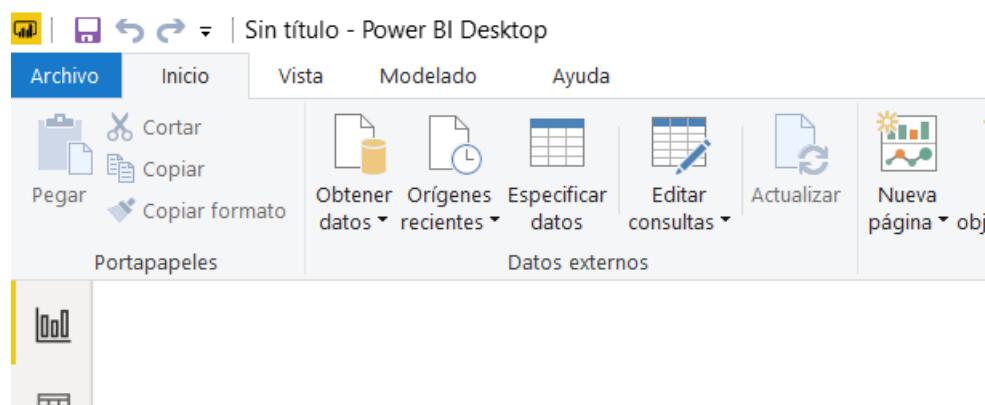
Al abrir Power BI, dependiendo de cómo lo tengamos configurado, nos aparecerá una pantalla de bienvenida desde la que directamente podemos establecer la conexión al origen de datos o abrir archivos recientes.



Pantalla de bienvenida en Power BI

Fuente: Elaboración propia

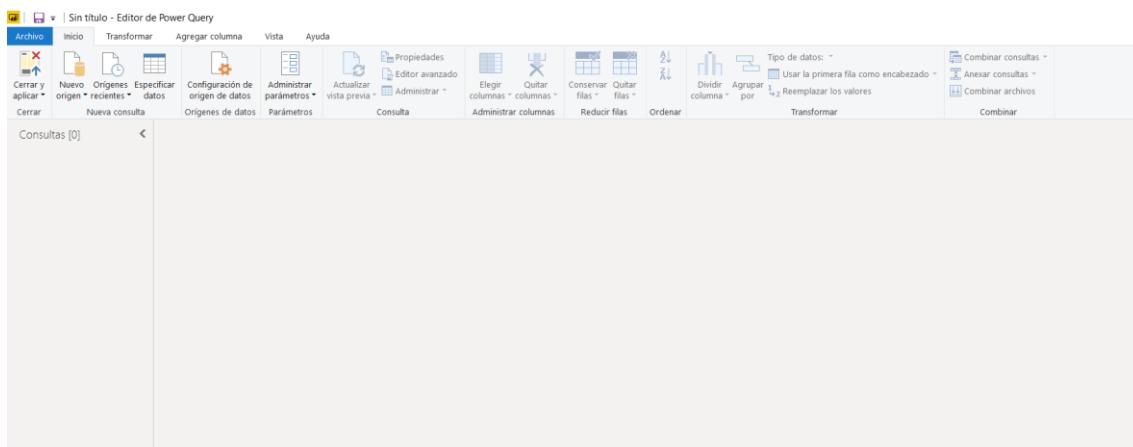
Desde esta pantalla, tan solo tendríamos que seleccionar la opción obtener datos para iniciar la conexión con el origen de datos. Si no se abre de forma predeterminada esta pantalla al abrir Power BI, tan solo tendremos que hacer clic en obtener datos en el menú Inicio. También podremos abrir la ventana del Power Query haciendo clic en el botón editar consultas.



Opción Obtener Datos en el menú Inicio en Power BI

Fuente: Elaboración propia

Si pinchamos en editar consultas se abrirá la ventana del Power Query. Para poder activar todas las opciones del Power Query necesitaremos conectarnos a un origen de datos.



Editor Power Query

Fuente: Elaboración propia

Cuando seleccionamos la opción obtener datos (ya sea desde la pantalla principal de Power BI o desde la pantalla de Power Query) vemos los conectores a orígenes externos de que dispone Power BI. Podemos conectarnos a archivos que tengamos tanto en local en nuestro ordenador como podría ser un archivo Excel o XML, como a servicios en línea, por ejemplo, Google Analytics o Mailchimp.

Siempre que nos conectemos a un origen de datos debemos modelar esos datos en Power Query para detectar y limpiar los posibles errores.

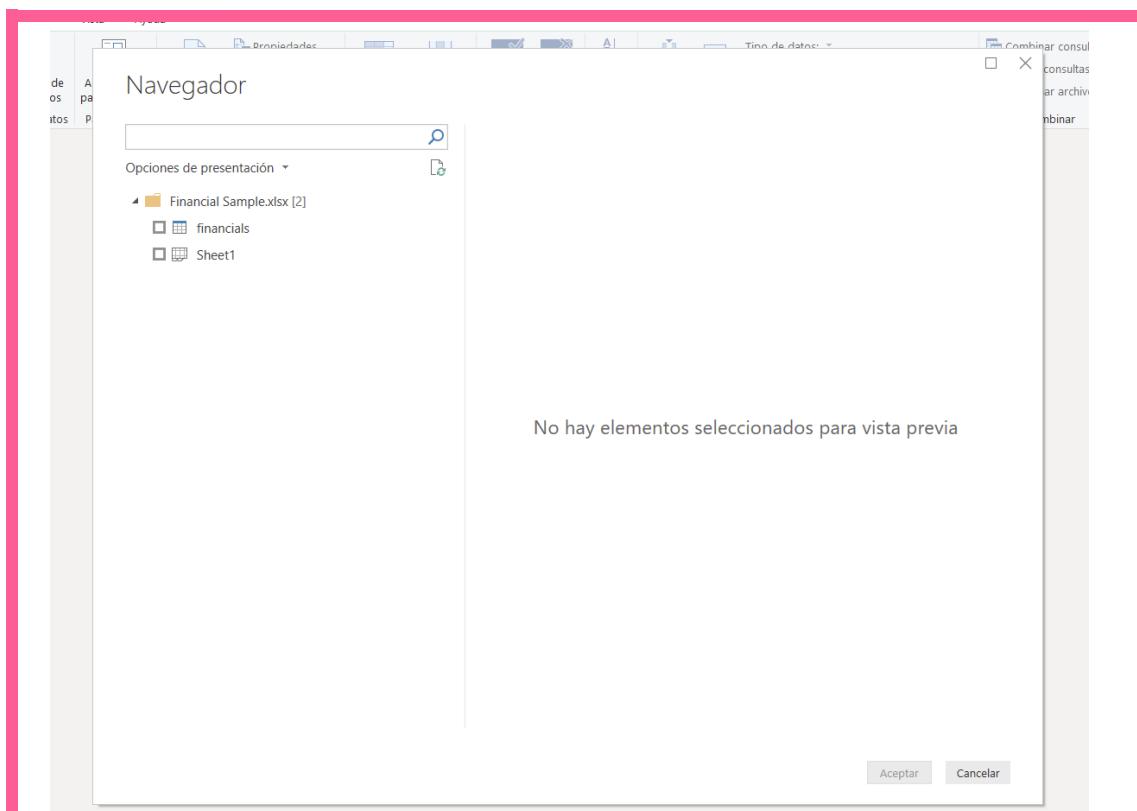


### EJEMPLO PRÁCTICO

Será muy común que tengamos que conectarnos a un archivo local para obtener los datos para nuestro modelo tabular.

Cuando nos conectamos a un archivo de Excel, al seleccionar el archivo, se mostrarán todas las hojas y tablas que contenga dicho archivo y tendremos que marcar aquellas que queramos importar a nuestro modelo de datos. Si no necesitamos todas las tablas u hojas, no debemos importarlas a Power Query, porque estaremos añadiendo información innecesaria al modelo.

Por ejemplo, en la siguiente pantalla vemos como al seleccionar un archivo de Excel para importar los datos, Power Query detecta que hay una hoja (que en esta imagen se llama Sheet1) y una tabla (que se llama Financials). Distinguiremos las hojas de las tablas porque aparecen representados con iconos diferentes. Si en el documento de Excel la hoja se llamase igual que la tabla, Power Query añadiría un número al final del nombre de la tabla para diferenciarlos, porque no se podrían importar dos orígenes de datos con el mismo nombre.



### Previsualización del contenido en Power BI

Fuente: Elaboración propia

Al seleccionar un elemento tendremos una vista previa de su contenido, lo que nos permitirá asegurarnos de que los datos que estamos importando son los que nos interesan.

## Navegador

The screenshot shows the Microsoft Power BI Navigator interface. On the left, there's a tree view of the file structure under 'Financial Sample.xlsx [2]'. The 'financials' sheet is selected, indicated by a yellow checkmark. Below the tree view is a search bar and a 'Presentación' dropdown menu. The main area displays a preview of the 'financials' data in a tabular format. The columns are labeled: Segment, Country, Product, Discount Band, and Unit. The data includes rows for Government, Midmarket, Channel Partners, and Enterprise segments across various countries like Canada, Germany, France, Mexico, and the United States. At the bottom right of the preview area are 'Aceptar' (Accept) and 'Cancelar' (Cancel) buttons.

Segment	Country	Product	Discount Band	Unit
Government	Canada	Carretera	None	
Government	Germany	Carretera	None	
Midmarket	France	Carretera	None	
Midmarket	Germany	Carretera	None	
Midmarket	Mexico	Carretera	None	
Government	Germany	Carretera	None	
Midmarket	Germany	Montana	None	
Channel Partners	Canada	Montana	None	
Government	France	Montana	None	
Channel Partners	Germany	Montana	None	
Midmarket	Mexico	Montana	None	
Enterprise	Canada	Montana	None	
Small Business	Mexico	Montana	None	
Government	Germany	Montana	None	
Enterprise	Canada	Montana	None	
Midmarket	United States of America	Montana	None	
Government	Canada	Paseo	None	
Midmarket	Mexico	Paseo	None	
Channel Partners	Canada	Paseo	None	
Government	Germany	Paseo	None	
Channel Partners	Germany	Paseo	None	
Government	Mexico	Paseo	None	
Midmarket	France	Paseo	None	

### Vista previa a la importación en Power BI

Fuente: Elaboración propia

Cuando el archivo de origen es un archivo de texto o CSV, tendremos que elegir de qué forma está codificado y cuál es el carácter delimitador para que Power Query pueda importar los datos al modelo tabular de la manera correcta.

The screenshot shows the 'Get Data' dialog in Power BI. The 'Origin' dropdown is set to '65001: Unicode (UTF-8)'. The 'Separator' dropdown is set to 'Comma'. The 'Data type detection' dropdown is set to 'Based on the first 200 rows'. The main area displays the CSV data with columns: Country, LinkedOrganizationShippingStreet, LinkedOrganizationShippingCity, LinkedOrganizationShippingState, and LinkedOrganizationShippingPostalCode. The data includes entries like 'C/ Pinto, 76 2º Planta' in Parla, Madrid, and 'C/ de la Ingeniería, 4' in Leganés, Madrid.

### Importación de un fichero CSV en Power BI

Fuente: Elaboración propia

Una vez que hemos seleccionado el origen de los datos, se cargan al modelo tabular y se activan todas las opciones de Power Query.

The screenshot shows the Power Query editor interface. The ribbon is open with several transformation tools like 'Validar', 'Error', 'Vacio', 'Categorizar', 'Reducir filas', 'Dividir', 'Agrupar', 'Reemplazar los valores', and 'Transformar'. The main area displays a table with columns: Segment, Country, Product, Discount Rate, Units Sold, Manufacturing Price, and Sale Price. The 'Segment' column contains values like 'Government', 'Midmarket', and 'Channel Partners'. The 'Country' column lists countries like Canada, Germany, France, Mexico, and United States of America. The table has 20 rows of data. On the right side, there are sections for 'Configuración de la consulta' (Nombre: financials), 'PROPIEDADES' (Nombre: financials), and 'PASOS APLICADOS' (Origen, Navegación).

### Pantalla de Power Query con todas las opciones activadas

Fuente: Elaboración propia



## VIDEOTUTORIAL

En el siguiente vídeo te explicamos como conectarte a un origen de datos local en formato xls, csv o txt.

<https://vimeo.com/user64513894/review/447101139/473fcf4f57>

A la izquierda de la pantalla tenemos el panel consultas, en el que se almacenan todos los elementos con los que trabajamos en Power Query, ya sean consultas, funciones, parámetros u otros elementos.

Si seleccionamos varias tablas u hojas, como resultado tendremos varias consultas, cuyo origen es el mismo libro Excel.

En el panel central tenemos una vista previa de los datos. Cada columna cuenta con un par de botones en la parte superior que se utilizan para definir el tipo de datos y para realizar opciones de filtrado.

A la derecha se encuentra el panel configuración de consulta, que contiene las propiedades y los pasos que se han aplicado a la consulta.

Cada una de las tareas que realizamos en una consulta constituye un paso. Siempre que importemos un nuevo origen de datos para crear una consulta, tendremos varios pasos generados por Power BI, por ejemplo:

- Un paso que muestra que se ha establecido la conexión con el origen. Si es un documento de Excel nos mostrará todo lo que estaba en el archivo de origen.

Primer paso de Power Query que muestra que se ha establecido la conexión con el origen.

Fuente: Elaboración propia

- Un paso que muestra que se han importado los datos al modelo.

The screenshot shows the Microsoft Power BI Data Editor interface. A table named "financials" is displayed with the following columns: Segment, Country, Product, Discount Band, Units Sold, Manufacturing Price, and Sale Price. The "Manufacturing Price" column is currently selected. The ribbon at the top has tabs for Archivo, Inicio, Transformar, Agregar columna, Vista, and Ayuda. On the right side, there is a "Configuración de la consulta" pane with sections for PROPIEDADES (Nombre: financieras) and PASOS APLICADOS (Origen: Navegación, Tipo: Navegación).

Segundo paso que muestra que se han importado los datos.

Fuente: Elaboración propia

- Un paso que muestra que Power Query ha analizado y modificado el tipo de datos para que sean manipulables.

This screenshot shows the Microsoft Power BI Data Editor after a transformation step. The "Manufacturing Price" column has been converted to an integer type ("Número entero"). The rest of the table structure remains the same as in the previous screenshot.

Tercer paso que muestra que se ha modificado el tipo de datos.

Fuente: Elaboración propia

Las consultas que tienen como origen una tabla, presentan tres pasos. Mientras que las consultas cuyo origen es una hoja, presentan un paso adicional, que nos sirve para convertir la primera fila de cada hoja en la cabecera de las columnas.

Power Query cuenta además con una banda de opciones con varias pestañas –Inicio, transformar, agregar columna, vista y ayuda–, donde se agrupan las funcionalidades disponibles para crear y configurar consultas, así como agregar, transformar y eliminar filas y columnas en la consulta actual.

Las transformaciones que hagamos con los datos en Power Query las podemos hacer a través de la interfaz de usuario o a través del editor avanzado que encontramos en el menú vista.

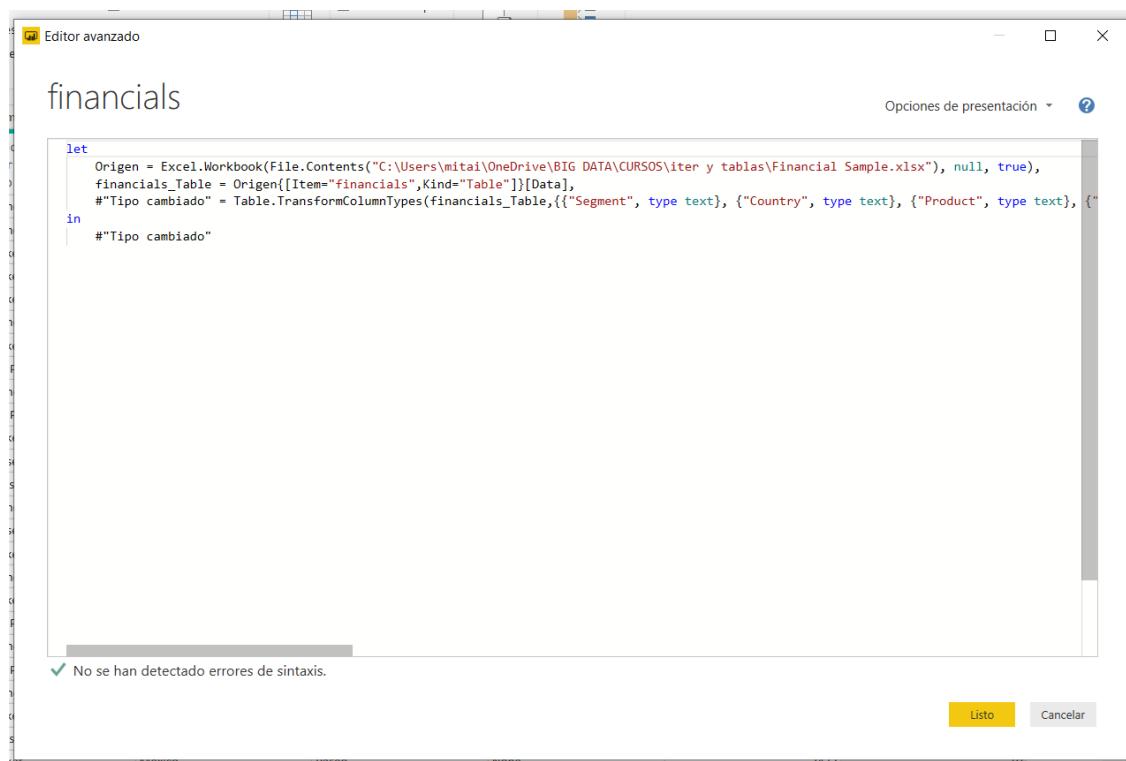
This screenshot shows the "Vista" tab of the Microsoft Power BI Data Editor. It includes several configuration options: Barra de fórmulas, Monoespacida, Distribución de columnas, Mostrar espacio en blanco, Perfil de columna, Calidad de columnas, Ir a columna, Permitir siempre, Editor avanzado, and Dependencias de la consulta. Below the ribbon, the "Consultas [1]" tab is selected, showing the table structure with columns: Segment, Country, Product, Discount Band, and Units Sold.

Menú Vista con el botón del Editor avanzado

Fuente: Elaboración propia

El editor avanzado nos permite introducir los cambios que queramos hacer en lenguaje M.

El lenguaje M es un lenguaje funcional de Power Query que distingue mayúsculas de minúsculas. Cualquier combinación o modificación de los datos se expresa en lenguaje M.



The screenshot shows the 'Editor avanzado' window in Excel. The title bar says 'Editor avanzado'. The main area contains the following M code:

```
let
    Origen = Excel.Workbook(File.Contents("C:\Users\mitai\OneDrive\BIG DATA\CURSOS\iter y tablas\Financial Sample.xlsx"), null, true),
    financials_Table = Origen{[Item="financials",Kind="Table"]}[Data],
    #"Tipo cambiado" = Table.TransformColumnTypes(financials_Table,{{"Segment", type text}, {"Country", type text}, {"Product", type text}, {"Type", type text}}),
in
    #"Tipo cambiado"
```

At the bottom left, there is a green checkmark icon followed by the text 'No se han detectado errores de sintaxis.' (No syntax errors detected). At the bottom right, there are 'Listo' (Ready) and 'Cancelar' (Cancel) buttons. The status bar at the bottom shows 'Editor avanzado' and 'Fuente: Elaboración propia'.

Editor avanzado de Power Query

Fuente: Elaboración propia



### ENLACE DE INTERÉS

En esta página de Microsoft puedes aprender más sobre el lenguaje M, que también puedes utilizar en Excel.

<https://docs.microsoft.com/es-es/powerquery-m/>

Los tres primeros pasos los crea el motor de Power BI. Los siguientes los accionamos nosotros cuando aplicamos transformaciones sobre las columnas.

De todas las columnas que hemos importado, si trabajamos a partir de una tabla de Excel, tal vez no todas sean necesarias, por lo que lo mejor será eliminar las que sobren.

Para ello vamos a la banda de opciones del menú inicio, grupo administrar columnas y pinchamos en el botón elegir columnas. Seleccionamos las columnas que queremos conservar y, cuando hayamos terminado, seleccionaremos entonces el botón aceptar.

The screenshot shows the 'Elegir columnas' (Select Columns) dialog box in Power Query. On the left, a list of columns is shown with checkboxes next to them. The columns selected for retention are: Segment, Country, Product, Units Sold, Manufacturing Price, Sale Price, and Gross Sales. Below this list is a preview of the data with these columns. At the bottom right of the dialog are 'Aceptar' (Accept) and 'Cancelar' (Cancel) buttons.

Selección de columnas en Power Query

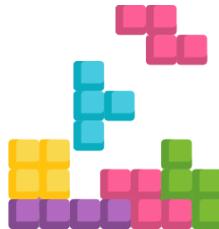
Fuente: Elaboración propia

Como vemos, a continuación, en el panel configuración de la consulta se ha creado un nuevo paso.

The screenshot shows the Power Query configuration pane. In the 'PASOS APlicados' (Applied Steps) section, there is a step named 'financials'. This indicates that the user has created a new step after selecting the columns. The main area shows the data with the columns 'Country', 'Product', 'Units Sold', 'Manufacturing Price', 'Sale Price', 'Gross Sales', and 'Discounts' selected for retention.

Nuevo paso creado en Power Query tras seleccionar las columnas

Fuente: Elaboración propia



## EJEMPLO PRÁCTICO

Con cierta frecuencia necesitaremos modificar el tipo de dato que ha asignado Power Query de forma automática, o hacer modificaciones de forma global en la columna.

Por ejemplo, si tenemos una fecha completa y solo nos interesa conservar el mes, porque todas las filas son del mismo año, hacemos clic sobre la columna en la que está la fecha y, desde la ficha transformación de la banda de opciones de Power Query, nos vamos a las transformaciones sobre columna de tipo fecha y hora.

En el botón fecha desplegamos el menú y seleccionamos la opción que más nos interese.

Menú desplegado en el botón fecha de Power Query

Fuente: Elaboración propia

Como resultado, añadiremos un nuevo paso y el tipo de dato se cambiará al tipo correspondiente. Es decir, que si seleccionamos que se transforme de forma que quede solo el nombre del mes, el tipo de dato de la columna dejará de ser tipo fecha y pasará a ser tipo texto.

The screenshot shows the Microsoft Power Query Editor interface. On the left, there's a table with columns labeled 'S' and 'Date'. The 'Date' column contains dates like '32370 enero', '26420 enero', '32670 junio', etc. To the right of the table is a 'Configuración de la consulta' (Query Configuration) pane. Under 'PROPIEDADES' (Properties), the 'Nombre' (Name) is set to 'financials'. Under 'PASOS APlicados' (Applied Steps), the step 'Nombre del mes extraído' (Month name extracted) is highlighted. The status bar at the bottom indicates 'Paso creado en el que se ha extraído el nombre del mes' (Step created where the month name was extracted).

Paso creado en el que se ha extraído el nombre del mes

Fuente: Elaboración propia

Si queremos realizar una consulta sobre documentos con otro tipo de formato que no sea necesariamente una tabla, como podría ser un documento en XML, Power Query es capaz de interpretar esa información y mostrarla en el modelo tabular. XML es un formato muy utilizado para el intercambio de información, de forma que es interesante saber cómo se realiza la conexión con estos archivos.

Por ejemplo, si tenemos un documento XML con la siguiente estructura:

```
<?xml version="1.0"?>
<catalogo>
<Libro id="1">
<autor>Gambardella, Matthew</autor>
<titulo>XML Developer's Guide</titulo>
<genero>Computer</genero>
<ventas>44,95</ventas>
<fecha>2000-10-01</fecha>
</Libro>
<Libro id="2">
<autor>Ralls, Kim</autor>
<titulo>Midnight Rain</titulo>
<genero>Fantasy</genero>
```

```

<ventas>5,95</ventas>
<fecha>2000-12-16</fecha>
</Libro>
</catalogo>

```

Cuando establecemos la conexión con el origen, en la vista previa, antes de cargar los datos, ya nos los mostrará como una tabla.

autor	titulo	genero
Gambardella, Matthew	XML Developer's Guide	Computer
Rails, Kim	Midnight Rain	Fantasy
Corets, Eva	Maeve Ascendant	Fantasy
Corets, Eva	Oberon's Legacy	Fantasy
Corets, Eva	The Sundered Grail	Fantasy
Randall, Cynthia	Lover Birds	Romance
Thurman, Paula	Splish Splash	Romance
Knorr, Stefan	Creepy Crawlies	Horror
Kress, Peter	Paradox Lost	Science Fiction
O'Brien, Tim	Microsoft .NET: The Programming Bible	Computer
O'Brien, Tim	MSXML3: A Comprehensive Guide	Computer
Galos, Mike	Visual Studio 7: A Comprehensive Guide	Computer

Vista previa de los datos en Power Query de un archivo en XML

Fuente: Elaboración propia

Cuando aceptamos e importamos los datos, Power Query seleccionará el tipo de dato adecuado para cada columna, como ya hemos explicado con anterioridad.

### Datos importados de un archivo XML

Fuente: Elaboración propia

XML es un lenguaje en el cual usar tildes puede generar problemas, por lo que, habitualmente, los desarrolladores no las utilizan, aunque pongan los nombres de los campos en español. En Power Query no tenemos ese tipo de problemas, así que, si queremos, podremos cambiar los nombres de las columnas y añadir las tildes.

Otro formato que se utiliza cada vez más para el intercambio de datos es JSON y, Power BI cuenta con un conector nativo para este tipo de archivos. Sin embargo, en este caso, es necesario añadir más pasos que cuando importamos datos de archivos XML.

Al obtener datos en formato JSON en Power Query tendremos que convertir los datos en una tabla. En primer lugar, tendremos un enlace a un objeto tipo lista y al abrir la lista veremos el grupo de registros almacenados. Cada registro es una fila de la lista y para previsualizar su contenido tan solo tenemos que hacer clic en la fila (pero no en el enlace al registro). De esta manera, en la parte inferior tendremos la visualización del contenido.

Una vez que tenemos seleccionado el registro, desde el menú transformar, seleccionamos la opción a la tabla. El resultado será una tabla con una columna en la que en cada fila tenemos el enlace a un registro. A continuación, para ver el contenido de los registros, haremos clic en el botón expandir columnas de la parte superior de la columna.

Archivo Inicio Transformar Agregar columna Vista Ayuda

Aprendizaje Documentación Vídeos de Soporte Blog de Comunidad Power BI para Ejemplos Galerías de la Envíe una id  
guiado aprendizaje técnico Acerca de Power BI desarrolladores desarrolladores comunidad una id

Ayuda

Comunidad

Consultas [4]

ABC Column1

	ABC Column1	
•	Válido	100 %
●	Error	0 %
●	Vacío	0 %
1	Record	
2	Record	
3	Record	
4	Record	
5	Record	
6	Record	
7	Record	
8	Record	
9	Record	
10	Record	

Datos importados de un archivo JSON

Fuente: Elaboración propia

Seleccionamos todas las columnas, y ya nos aparecerá una tabla con todo el contenido del archivo JSON.



### VIDEOTUTORIAL

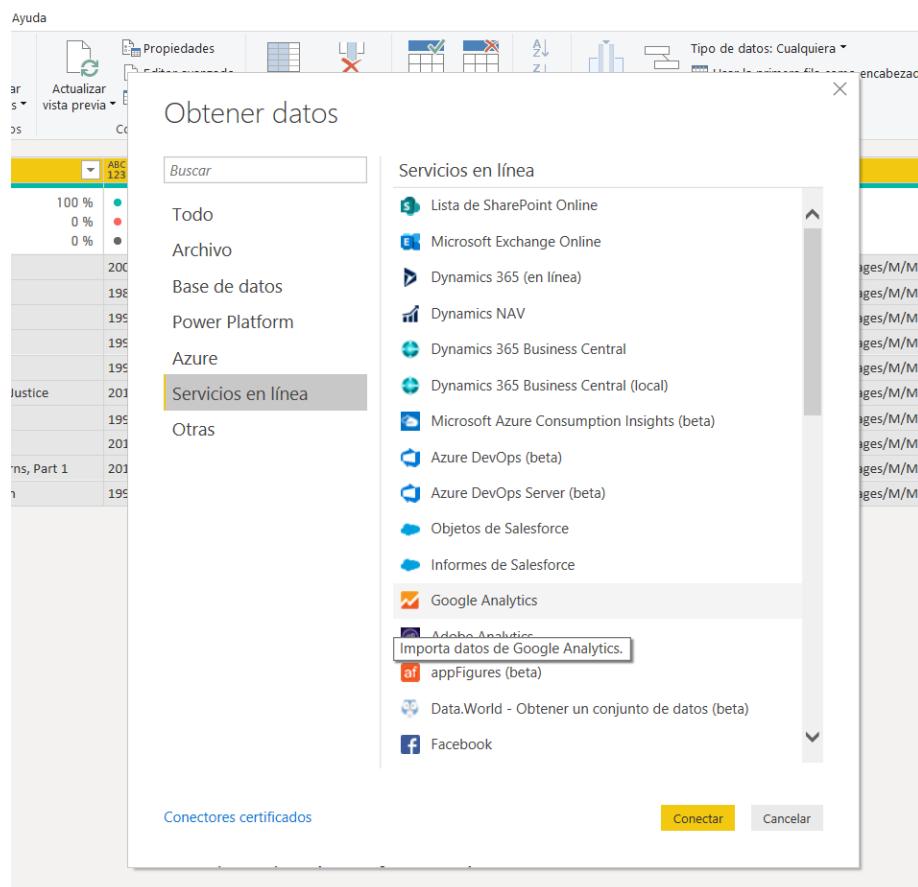
En este vídeo podemos ver de forma breve como es el proceso de importar archivos XML y JSON.

<https://vimeo.com/user64513894/review/447102324/7ea9d9c8bf>

Cuando lo que queremos es acceder a un servicio en línea para obtener nuestros datos, también tenemos conectores nativos que nos permiten integrar en nuestro modelo de datos información procedente de las principales plataformas.

Por ejemplo, podremos importar datos de Google Analytics. Al conectarnos a Analytics tendremos que introducir nuestro nombre de usuario y contraseña, porque de lo contrario, Power BI no podría acceder a la información.

Google Analytics es un servicio gratuito de Google que muestra un gran conjunto de datos de una página web que nos permiten monitorizar quién la visita, desde dónde llegan a nuestra página, cuánto tiempo dura cada sesión, etc. Además, es posible integrar Analytics con otras herramientas de Google como AdWords y AdSense.



### Obtener datos de servicios en línea con Power Query

Fuente: Elaboración propia

Al conectarnos a Analytics tendremos que introducir nuestro nombre de usuario y contraseña porque de lo contrario, Power BI no podría acceder a la información.



Pantalla para establecer la conexión con Google Analytics

Fuente: Elaboración propia

Una vez que hemos iniciado la sesión, tendremos acceso a toda la información que tenemos disponible en Google Analytics y deberemos seleccionar aquella que consideramos necesaria en nuestro modelo de datos. Por ejemplo, en la siguiente imagen podemos ver que solo nos interesa seleccionar algunos campos relacionados con nuestra campaña de Adwords.

Google Ads Campaign ID	Clicks	Cost per Conversion	CPC	CTR
(not set)	0	0	0	0
1071218774	33648	0,278489889	0,148729889	0,6
1708273094	6742	0,659347908	0,172025062	0,3
1943361288	0	0	0	0
1976497321	0	0	0	0
1976883555	0	0	0	0
8189321439	1609	0,873606968	0,18840374	0,1
8670423105	301	16,61666667	2,815448505	0,1
8768915267	0	0	0	0
8769849340	0	0	0	0
8919415144	1750	0,913257445	0,174823568	0,1
9073282259	121	3,966153846	0,426115702	0,1
9310067162	51	410,79	8,054705882	0,1
9506092453	0	0	0	0

Selección de datos de Adwords dentro de Google Analytics

Fuente: Elaboración propia

Una vez seleccionados los datos y aceptamos, los datos se importarán como una nueva consulta. Por defecto, la consulta se llamará «All Web Site Data», pero podremos renombrarla.

	Clicks	\$ Cost per Conversion	\$ CPC	CTR	Impressions	% New
1 (not set)	0	0	0	0	0	0
2 1071218774	33648	0,278499889	0,148729889	0,036331127	926097	
3 1708273094	6742	0,659347908	0,172025062	0,162783398	41417	
4 1943361388	0	0	0	0	0	
5 1976497321	0	0	0	0	0	
6 1976888555	0	0	0	0	0	
7 8189321439	1609	0,873606968	0,18840174	0,182488375	8817	
8 8670423105	301	16,61666667	2,815448505	0,178683015	1685	
9 8768915267	0	0	0	0	0	
10 8769849340	0	0	0	0	0	
11 8919415144	1750	0,913257445	0,174825568	0,162007036	10802	
12 9073282259	121	2,966153846	0,426115702	0,233590734	518	
13 9310067162	51	410,79	8,054705882	0,129770992	393	
14 9506092453	0	0	0	0	0	

Datos importados desde Google Analytics

Fuente: Elaboración propia

Power Query nos da también la opción de conectarnos a cualquier página web. Esto nos permitirá obtener datos de cualquier tipo de fuente de información. Power BI hace un excelente trabajo de detección de tablas en páginas y sitios web, y su conversión a consultas de Power Query. Cuando creamos una consulta a partir del contenido de una página web, después tendremos que preparar los datos para nuestro modelo: eliminar las filas, las columnas innecesarias y mejorar la estructura de la tabla.



### VIDEOTUTORIAL

En este videotutorial podrás ver cómo podemos importar datos de Wikipedia o Google Analytics a nuestro modelo de datos.

<https://vimeo.com/user64513894/review/447103575/4a33bec572>

A veces, nos vamos a encontrar que los datos que queremos importar en nuestro modelo se encuentran en un formato para el que no hay un conector nativo en Power BI. Si no existe ese conector nativo no podremos enlazar al origen de los datos y, por lo tanto, no podremos actualizar las consultas. En esos casos, lo que haremos es introducir los datos manualmente en Power Query.

Por ejemplo, supongamos que tenemos la siguiente tabla en documento de texto y queremos incluir estos datos en nuestro modelo de Power BI:

<b>Orden</b>	<b>CDU/ISBN</b>	<b>Descripción CDU</b>
1	0	Generalidades
2	001	Ciencia y conocimiento en general. Organización del trabajo intelectual. Divulgación de la ciencia.
3	001.83	Cooperación intelectual (técnica y científica). Teoría.
4	002	Documentación en general. Producción documental. Centros de documentación.
5	003	Escritura. Sistemas de escritura. Signos y símbolos. Representaciones gráficas en general.
6	004	Ciencia y tecnología de los ordenadores. Informática.
7	004.3	"Hardware". Componentes físicos del ordenador, máquinas.
8	004.4	"Software". Equipo lógico, componentes lógicos, programas.
9	004.6	Datos. Sistemas de gestión de bases de datos.
10	004.7	Comunicación informática. Redes de ordenadores.
11	005	Estudios teóricos de organización en general. Metodología. Análisis, síntesis, clasificación y taxonomía. Sistematización en general.
12	006	Normalización en general y normas.
13	006.9	Cronología.
14	006.91	Metrología. Pesos y medidas en general.
15	007	Teoría de la comunicación y del control en general. Cibernética.
16	008	Civilización, progreso y cultura en general.
17	009	Humanidades.
18	011	Bibliografías universales.
19	016	Bibliografías especiales.
20	017	Catálogos en general.
21	019	Catálogos especiales.

Desde la ficha Inicio del editor de consultas de Power Query pinchamos sobre la opción Especificar datos, del grupo Nueva consulta. Power Query abre una ventana vacía que está preparada para que creemos una tabla manualmente desde cero.

Screenshot of the Microsoft Power Query interface showing the 'Crear tabla' (Create Table) dialog. The interface includes a ribbon with 'Archivo', 'Inicio', 'Transformar', 'Agregar columna', 'Vista', and 'Ayuda'. A left sidebar lists 'Consultas [6]' with items like 'Ventas', 'Cuentas', 'Libro', 'Batman', 'Analytics', and 'Población extranjera en...'. The main area shows a table with one column labeled 'Columna1' containing 27 rows of data. A legend indicates three status levels: 'Válido' (green dot), 'Error' (red dot), and 'Vacio' (black dot). The right side shows a preview pane for 'L2 Discounts' with similar status indicators. A bottom bar includes 'Nombre:' (Name:), 'Aceptar' (Accept), and 'Cancelar' (Cancel).

Columna1
1 Válido
2 Error
3 Vacio
4 Válido
5 Válido
6 Válido
7 Válido
8 Válido
9 Válido
10 Válido
11 Válido
12 Válido
13 Válido
14 Válido
15 Válido
16 Válido
17 Válido
18 Válido
19 Válido
20 Válido
21 Válido
22 Válido
23 Válido
24 Válido
25 Válido
26 Válido
27 Válido

### Especificación de datos de forma manual en Power Query

Fuente: Elaboración propia

Una vez que estamos en esa pantalla resultaría tan sencillo como copiar los datos que queremos agregar al modelo y pegarlos. Si no podemos copiar y pegar (porque, por ejemplo, los datos están en una imagen) podemos añadirlos manualmente.

Crear tabla

**La primera fila de datos que pegó ascendió a los encabezados de columna.**

	Orden	CDU/ISBN	Descripción ...	*
1	1	0	Generalidades	
2	2	001	Ciencia y cono...	
3	3	001.83	Cooperación in...	
4	4	002	Documentació...	
5	5	003	Escritura. Siste...	
6	6	004	Ciencia y tecno...	
7	7	004.3	Hardware. Com...	
8	8	004.4	Software. Equi...	
9	9	004.6	Datos. Sistema...	
10	10	004.7	Comunicación i...	
11	11	005	Estudios teóric...	
12	12	006	Normalización ...	
13	13	006.9	Cronología.	
14	14	006.91	Metrología. Pe...	
15	15	007	Teoría de la co...	
16	16	008	Civilización, pr...	
17	17	009	Humanidades.	
18	18	011	Bibliografías un...	
19	19	016	Bibliografías es...	
20	20	017	Catálogos en g...	
21	21	019	Catálogos espec...	
*				

Nombre: Clasificación

Aceptar Cancelar

### Creación de tabla en Power Query

Fuente: Elaboración propia

Por defecto, la primera fila se convertirá en la fila de encabezados y la tabla se llamará Tabla, pero podemos cambiar el nombre en la parte inferior de la pantalla. Una vez añadidos los datos, al hacer clic en Aceptar aparecerá la tabla como otra consulta de nuestro modelo de datos. Si necesitásemos actualizar estos datos, nos vamos al paso Origen, seleccionamos la ruedecita y volvemos a tener la misma tabla que habíamos creado antes con opciones de configuración.

Consultas [7]

	Orden	CDU/ISBN	Descripción CDU
1	1	0	Generalidades
2	2	001	Ciencia y conocimiento en general. Organización del trabajo intelectual...
3	3	001.83	Cooperación intelectual (técnica y científica). Teoría...
4	4	002	Documentación en general. Producción documental. Centros de docu...
5	5	003	Escrutina. Sistemas de escritura. Signos y símbolos. Representaciones g...
6	6	004	Ciencia y tecnología de los ordenadores. Informática.
7	7	004.3	Hardware. Componentes físicos del ordenador, máquinas.
8	8	004.4	Software. Equipo lógico, componentes lógicos, programas.
9	9	004.6	Datos. Sistemas de gestión de bases de datos.
10	10	004.7	Comunicación informática. Redes de ordenadores.
11	11	005	Estudios teóricos de organización en general. Metodología. Análisis, si...
12	12	006	Normalización en general y normas.
13	13	006.9	Cronología.
14	14	006.91	Metrología. Pesos y medidas en general.
15	15	007	Teoría de la comunicación y del control en general. Cibernetica.
16	16	008	Civilización, progreso y cultura en general.
17	17	009	Humanidades.
18	18	011	Bibliografías universales.
19	19	016	Bibliografías especiales.
20	20	017	Catálogos en general.
21	21	019	Catálogos especiales.

Configuración de la consulta

PROPIEDADES

Nombre: Clasificación

TODAS LAS PROPIEDADES

PASOS APLICADOS

Origen

Tipo cambiado

### Selección del paso origen para modificar la tabla creada en Power Query

Fuente: Elaboración propia

## 4.5 Transformar las consultas en Power BI

En Power BI para poder trabajar con las consultas debemos asegurarnos de que los tipos de datos de cada columna son los adecuados, porque de lo contrario no podremos utilizarlos. Por ejemplo, si no hemos definido una columna como tipo de datos Fecha, no podremos seleccionar los días a los que vamos a aplicar la Consulta.

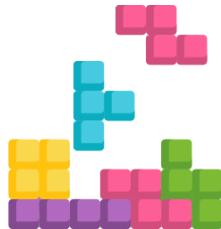
En Power Query podemos identificar los tipos de datos en la esquina superior izquierda de cada columna. Cuando haces clic en ese botón se desplegarán los tipos de datos disponibles. El ícono que aparece en el botón nos indica el tipo de dato de esa columna.

The screenshot shows the Power Query Editor interface. On the left, there is a preview pane with 13 rows of data. The first column is labeled 'RecordId' and contains numerical values from 1 to 13. The second column is labeled 'FirstName' and contains names like 'achraf'. The third column is labeled 'LastName' and contains names like 'amazian'. Above the preview, there is a status bar showing '123 RecordId' and a progress bar at 100%. To the right of the preview, there is a column header for 'FirstName' with a dropdown arrow. A context menu is open over this arrow, listing various data types: 'Número decimal', 'Número decimal fijo', 'Número entero', 'Porcentaje', 'Fecha/Hora', 'Fecha', 'Hora', 'Fecha/Hora/Zona horaria', 'Duración', 'Texto', 'Verdadero/Falso', 'Binario', and 'Usar configuración regional...'. The 'Número decimal' option is currently selected. The status bar also shows '100 %' and '0 %'.

Selección del tipo de dato en Power Query

Fuente: Elaboración propia

Power Query analiza el contenido de las columnas y selecciona el tipo de datos que considera adecuado, pero hay que comprobar que sea el tipo correcto.



## EJEMPLO PRÁCTICO

En la siguiente imagen vemos como la columna Fecha de nacimiento el tipo de datos que ha asignado Power Query es Texto, cuando debería ser Fecha/Hora.

The screenshot shows the Power BI Data Editor interface. A context menu is open over the 'Fecha de nacimiento' column header, specifically the dropdown menu for data types. The menu title is 'Cambiar tipo de dato' (Change data type). It lists three options: 'Válido' (Valid), 'Error' (Error), and 'Vacío' (Empty), with 'Válido' selected. Below the menu, the 'Fecha de nacimiento' column is highlighted in yellow, and its current data type is shown as 'Text'. The table contains 18 rows of data, mostly names and dates. Some dates are correctly formatted as 'dd/MM/yyyy' (e.g., '10/06/2000'), while others are in 'MM/dd/yyyy' (e.g., '06/10/2000') or even plain text ('-'). The 'Tienda' column shows locations like 'Torrejón', 'Leganés', and 'Alvarez Caletrio'.

A <sup>B</sup> <sub>C</sub> LastName	A <sup>B</sup> <sub>C</sub> Fecha de nacimiento	A <sup>B</sup> <sub>C</sub> Tienda
• Válido	100 %	• Válido
• Error	0 %	• Error
• Vacío	0 %	• Vacío
Álvarez Caletría		
-		Leganés
-		
Khotaibi	10/6/2000 12:00:00 AM	Torrejón
El Amrani	8/21/1996 12:00:00 AM	
El Meskini	1/20/2002 12:00:00 AM	Leganés
Ghesmoune	7/10/1998 12:00:00 AM	Torrejón
Silva Salazar		
Martínez Rovelo	9/17/2000 12:00:00 AM	Leganés
Leon	3/8/1998 12:00:00 AM	Torrejón
Romero Rolle	4/20/1997 12:00:00 AM	
-		
amazian	10/28/1992 12:00:00 AM	
Sackl	3/4/2018 12:00:00 AM	
Tayfi	4/26/1995 12:00:00 AM	
Arras	7/10/2000 12:00:00 AM	Leganés
Beranilla	4/21/2001 12:00:00 AM	Leganés

Columna con el tipo de dato mal seleccionado

Fuente: Elaboración propia

Este error se ha producido porque la configuración regional del archivo de origen de los datos es diferente a la configuración regional que tenemos configurada en Power BI. En el archivo de origen se utiliza una configuración de fecha/hora americana de forma que, al escribir la fecha, se indica primero el mes, después el día y por último el año. Sin embargo, si tenemos Power BI con una configuración española, Power Query reconocerá las fechas cuando se escribe primero el día, después el mes y por último el año. Al no reconocer los datos como fecha, automáticamente los deja como texto. Esto lo solucionaremos seleccionando la configuración regional adecuada.



## VIDEOTUTORIAL

En este vídeo podemos ver cómo crear nuestra propia tabla y modificar los tipos de datos cuando no son los adecuados:

<https://vimeo.com/user64513894/review/447104672/e7c188e2a6>

La configuración regional también es muy importante cuando trabajamos con datos de tipo numérico porque a veces no se interpreta de la manera correcta los decimales. En España el punto se utiliza para separar estéticamente los miles, mientras la coma se emplea para los valores decimales, pero esto no ocurre en todos los países, de forma que también tendremos que revisar que la configuración regional de los datos de origen coincide con nuestra configuración local.

Es muy importante para poder emitir los informes en Power BI que los tipos de datos queden bien definidos antes de cargar los datos al modelo tabular. Al conectarnos al origen de datos, Power BI navega por los datos y, en ocasiones no quedan definidos los tipos de datos. Esta información la tenemos si vamos a la esquina superior izquierda de cada columna y vemos que se indica **ABC 123**. Esta indicación corresponde al tipo de datos cualquiera. Un tipo de datos cualquiera no debe pasar al modelo tabular, por lo que tendremos que asignar el tipo que corresponda.

Una vez que creamos la consulta debemos limpiar las filas y columnas para eliminar los elementos innecesarios.

Lo primero que tendremos que hacer es seleccionar las columnas necesarias porque con frecuencia en la fuente de los datos nos encontraremos muchas columnas innecesarias. Esto lo haremos con la opción elegir columnas del grupo Administrar columnas.

Screenshot of Microsoft Power BI Data Editor showing a query named "Consultas [8]". The table has columns: RecordId, FirstName, LastName, and Fecha de nacimiento. A context menu is open over the LastName column, with the "Elegir columnas" option highlighted.

	RecordId	FirstName	LastName	Fecha de nacimiento
1	46440661	Aarón	Álvarez Caletrío	10/6/2000 12:00:00 AM
2	46441119	Abdelilah	-	
3	46845829	Abdelilah	-	
4	47418636	Abdelilah	Khotaihi	10/6/2000 12:00:00 AM

### Selección de columnas

Fuente: Elaboración propia

Si cometemos un error y seleccionamos menos columnas de las necesarias, tan solo tendremos que hacer clic en el paso Otras columnas quitadas y nos aparecerá el listado de columnas disponibles en el origen de los datos.

Screenshot of the "Elegir columnas" (Select Columns) dialog box. It shows a list of columns from the data source, including RecordId, FirstName, LastName, and several other fields like Salutation, Organization, Title, Phone, Fax, Email, Website, Street, City, State, PostalCode, Country, Industry, EmployeeCount, and Rating. The "Otras columnas quitadas" (Other columns removed) section is highlighted, listing the columns that were removed: FirstName, LastName, Organization, Title, Phone, Fax, Email, Website, Street, City, State, PostalCode, Country, Industry, EmployeeCount, and Rating.

### Modificación de columnas seleccionadas

Fuente: Elaboración propia

Debemos evitar cargar en el modelo tabular columnas calculadas porque son datos que podemos obtener a partir de otras columnas. Por ejemplo, en la siguiente imagen podemos ver que en la columna Diferencia se está calculando en qué porcentaje ha variado la población extranjera en España por países de procedencia entre 2011 y 2017. Si tenemos la población extranjera de los dos años podremos calcular ese valor, de forma que introducir esa columna en el modelo resulta redundante y debemos quitarla.

	A <sup>B</sup> <sub>C</sub> Puesto	A <sup>B</sup> <sub>C</sub> País[23]	A <sup>B</sup> <sub>C</sub> 2017[24]	A <sup>B</sup> <sub>C</sub> 2011[25]	A <sup>B</sup> <sub>C</sub> Diferencia			
	● Válido ● Error ● Vacío	94 % 0 % 6 %	● Válido ● Error ● Vacío	100 % 0 % 0 %	● Válido ● Error ● Vacío	100 % 0 % 0 %	● Válido ● Error ● Vacío	100 % 0 % 0 %
1	Puesto	País	2017	2011	Diferencia			
2	1	Marruecos	769.050	773.895	-0,6%			
3	2	Rumania	673.017	865.707	-22,3%			
4	3	Reino Unido	240.934	391.194	-38,4%			
5	4	China	215.748	167.132	+29,1%			
6	5	Italia	206.066	187.993	+9,6%			
7	6	Colombia	165.608	273.176	-39,4%			
8	7	Ecuador	135.045	360.710	-62,6%			
9	8	Bulgaria	123.730	172.926	-28,4%			
10	9	Alemania	110.852	195.987	-43,4%			
11	10	Ucrania	106.823	86.316	+23,8%			
12	11	Bolivia	99.220	199.080	-50,2%			
13	12	Francia	98.558	122.503	-19,5%			
14	13	Venezuela	95.474	59.805	+59,6%			
15	14	Portugal	89.005	140.824	-36,8%			
16	15	Pakistán	82.738	70.165	+17,9%			
17		null Total	4.572.807	5.751.487	-20,5%			

Ejemplo de modelo tabular en el que hay que hacer correcciones para poder manipular los datos

Fuente: Elaboración propia

En la tabla de la imagen anterior vemos que los tipos de datos de todas las columnas se han cargado como si fuesen datos de texto, aunque podemos observar que en varias columnas son datos numéricos. Es fundamental prestar atención a los tipos de datos porque si queremos realizar cálculos con los datos de las columnas, si no son datos numéricos no podremos hacerlo. Por lo tanto, las columnas de la tabla anterior en nuestro modelo de datos deberían figurar de la siguiente forma:

	A <sup>B</sup> <sub>C</sub> País[23]	1 <sup>2</sup> <sub>3</sub> 2017[24]	1 <sup>2</sup> <sub>3</sub> 2011[25]
	<ul style="list-style-type: none"> <li>● Válido 100 %</li> <li>● Error 0 %</li> <li>● Vacío 0 %</li> </ul>	<ul style="list-style-type: none"> <li>● Válido - %</li> <li>● Error 5 %</li> <li>● Vacío - %</li> </ul>	<ul style="list-style-type: none"> <li>● Válido - %</li> <li>● Error 5 %</li> <li>● Vacío - %</li> </ul>
1	País	Error	Error
2	Marruecos	769050	773895
3	Rumania	673017	865707
4	Reino Unido	240934	391194
5	China	215748	167132
6	Italia	206066	187993
7	Colombia	165608	273176
8	Ecuador	135045	360710
9	Bulgaria	123730	172926
10	Alemania	110852	195987
11	Ucrania	106823	86316
12	Bolivia	99220	199080
13	Francia	98558	122503
14	Venezuela	95474	59805
15	Portugal	89005	140824
16	Pakistán	82738	70165
17	Total	4572807	5751487

Modelo tabular con los tipos de datos adecuados y sin columnas calculadas

Fuente: Elaboración propia

Al transformar las consultas en Power Query, no solo debemos hacer cambios en las columnas, sino también en las filas. Si seguimos con el mismo ejemplo, en la tabla anterior podemos ver que en la primera fila aparece la misma información que en el encabezado y que en las columnas 2017 y 2011 se muestra como un error la primera fila porque hemos marcado el tipo de dato a número entero.

Para solucionarlo, debemos ir a la ficha Transformar del menú de opciones de Power Query y, en el grupo Tabla, tenemos un botón que pone Usar la primera fila como encabezado.

Archivo Inicio Transformar Agregar columna Vista Herramientas Ayuda

Agrupar por Usar la primera fila como encabezado Transponer Invertir filas Contar filas Cambiar nombre Reemplazar los valores Rellenar Columna dinámica Anular dinamización de columnas Mover Convertir en lista

Usar la primera fila como encabezado Consultas [8] País[25] 123 2017[24] 123 2011[25] Cualquier columna

	Válido	100 %	Válido	- %	Válido	- %
	Error	0 %	Error	5 %	Error	5 %
	Vacio	0 %	Vacio	- %	Vacio	- %
1	País	Error		Error		
2	Marruecos		769050		773895	
3	Rumania		673017		865707	
4	Reino Unido		240934		391194	
5	China		215748		167132	
6	Italia		206066		187993	

### Selección de la primera fila como encabezado

Fuente: Elaboración propia

Después de hacer clic, tendremos el siguiente resultado en el que ya no tendremos errores ni la primera fila con el mismo contenido que el encabezado:

Archivo Inicio Transformar Agregar columna Vista Herramientas Ayuda

Agrupar por Usar la primera fila como encabezado Transponer Invertir filas Contar filas Cambiar nombre Reemplazar los valores Rellenar Columna dinámica Anular dinamización de columnas Mover Convertir en lista

Tabla Consultas [8] País[25] 123 2017 123 2011 Cualquier columna

	Válido	100 %	Válido	100 %	Válido	100 %
	Error	0 %	Error	0 %	Error	0 %
	Vacio	0 %	Vacio	0 %	Vacio	0 %
1	Marruecos		769050		773895	
2	Rumania		673017		865707	
3	Reino Unido		240934		391194	
4	China		215748		167132	
5	Italia		206066		187993	
6	Colombia		165608		273176	
7	Ecuador		135045		360710	
8	Bulgaria		123730		172926	
9	Alemania		110852		195987	
10	Ucrania		106823		86316	
11	Bolivia		99220		199080	
12	Francia		98558		122503	
13	Venezuela		95474		59805	
14	Portugal		89005		140824	
15	Pakistán		82738		70165	
16	Total		4572807		5751487	

### Resultado de aplicar la primera fila como encabezado

Fuente: Elaboración propia

Podemos ver que en la última fila de la tabla tenemos también una fila con el total. Al igual que no debemos incorporar al modelo columnas calculada, las filas calculadas también debemos eliminarlas.

Esto podemos hacerlo desde el grupo de opciones Reducir filas en la ficha Inicio.

The screenshot shows the Power BI desktop interface with the 'Inicio' ribbon selected. In the 'Administrador de columnas' group, the 'Reducir filas' button is highlighted with a blue arrow. A callout box labeled 'Opción para quitar filas' is positioned over this button. The main area displays a table titled 'País' with data for various countries and their populations in 2017 and 2011.

	123	2017	123	2011	
Válido	100 %	Válido	100 %	Válido	100 %
Error	0 %	Error	0 %	Error	0 %
Vacio	0 %	Vacio	0 %	Vacio	0 %
1 Marruecos		769050		773895	
2 Rumania		673017		865707	
3 Reino Unido		240934		391194	
4 China		215748		167132	
5 Italia		206066		187993	
6 Colombia		165608		273176	
7 Ecuador		135045		360710	
8 Bulgaria		123730		172926	
9 Alemania		110852		195987	
10 Ucrania		106823		86316	
11 Bolivia		99220		199080	
12 Francia		98558		122503	
13 Venezuela		95474		59805	
14 Portugal		89005		140824	
15 Pakistán		82738		70165	
16 Total		4572807		5751487	

#### Opción para quitar filas

Fuente: Elaboración propia

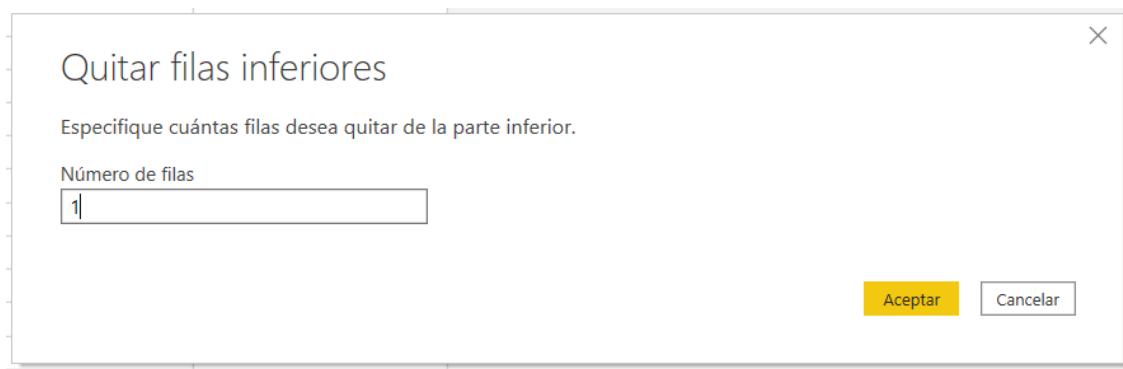
Seleccionamos la opción Quitar filas inferiores:

The screenshot shows the 'Quitar filas' dropdown menu open. The 'Quitar filas inferiores' option is highlighted with a blue arrow. The menu also includes other options like 'Quitar filas superiores', 'Quitar filas alternas', 'Quitar duplicados', 'Quitar filas en blanco', and 'Quitar errores'.

#### Detalle de la opción para quitar filas

Fuente: Elaboración propia

E indicamos que queremos quitar una fila.



Ventana para seleccionar el número de filas a quitar

Fuente: Elaboración propia

El resultado que tendremos es una tabla sin errores ni filas ni columnas con valores calculados en la que los tipos de datos son los correctos.

	A <sup>B</sup> <sub>C</sub> País	1 <sup>2</sup> <sub>3</sub> 2017	1 <sup>2</sup> <sub>3</sub> 2011
	<ul style="list-style-type: none"> <li>● Válido 100 %</li> <li>● Error 0 %</li> <li>● Vacío 0 %</li> </ul>	<ul style="list-style-type: none"> <li>● Válido 100 %</li> <li>● Error 0 %</li> <li>● Vacío 0 %</li> </ul>	<ul style="list-style-type: none"> <li>● Válido 100 %</li> <li>● Error 0 %</li> <li>● Vacío 0 %</li> </ul>
1	Marruecos	769050	773895
2	Rumania	673017	865707
3	Reino Unido	240934	391194
4	China	215748	167132
5	Italia	206066	187993
6	Colombia	165608	273176
7	Ecuador	135045	360710
8	Bulgaria	123730	172926
9	Alemania	110852	195987
10	Ucrania	106823	86316
11	Bolivia	99220	199080
12	Francia	98558	122503
13	Venezuela	95474	59805
14	Portugal	89005	140824
15	Pakistán	82738	70165

Tabla sin errores adecuada para el modelo tabular

Fuente: Elaboración propia

También desde la esquina superior izquierda de la tabla, tenemos la opción que nos permite quitar las últimas filas de la tabla.

Screenshot of Microsoft Power BI showing a context menu open over a table. The menu options include:

- Copiar la tabla completa
- Usar la primera fila como encabezado
- Agregar columna personalizada...
- Agregar columna a partir de los ejemplos...
- Invocar función personalizada...
- Agregar una columna condicional...
- Agregar columna de índice
- Elegir columnas...
- Conservar filas superiores...
- Conservar filas inferiores...
- Conservar intervalo de filas...
- Conservar duplicados
- Conservar errores
- Quitar filas superiores...
- Quitar filas inferiores...**
- Quitar filas alternas...
- Quitar duplicados
- Quitar errores
- Combinar consultas...
- Anexar consultas...

123	2017	123	2011
100 %	● Válido	100 %	
0 %	● Error	0 %	
0 %	● Vacío	0 %	
769050		773895	
673017		865707	
240934		391194	
215748		167132	
206066		187993	
165608		273176	
135045		360710	
123730		172926	
110852		195987	
106823		86316	
99220		199080	
98558		122503	
95474		59805	
89005		140824	
82738		70165	
4572807		5751487	

Opciones disponibles en el menú de la esquina superior izquierda de cada columna  
 Fuente: Elaboración propia



**VIDEOTUTORIAL**

En este vídeo podemos ver cómo realizar cambios en las filas de Power Query:

<https://vimeo.com/user64513894/review/447105731/98e310177f>

Es posible aplicar filtros a las columnas si no necesitamos incluir en el modelo todos los datos de la tabla. Por ejemplo, si solo necesitamos los valores que están por encima de una cantidad o los datos de personas procedentes de un lugar, etc., podremos aplicar filtros de texto o numéricos según corresponda. Los filtros disponibles variarán en función del tipo de dato almacenado en la columna y se activan en el botón de la esquina superior derecha de cada columna.

The screenshot shows the Microsoft Power Query Editor interface. On the left, there's a table with columns labeled 'País' and 'Número'. The 'Número' column has 15 entries ranging from 1 to 15, each associated with a country name. A context menu is open over the 'Número' column, specifically over the header cell '123 2017'. The menu includes options for sorting ('Orden ascendente', 'Orden descendente'), deleting ('Borrar orden', 'Borrar filtro', 'Quitar vacíos'), and filtering ('Filtros de número'). A sub-menu for 'Filtros de número' is expanded, showing various filter conditions like 'Es igual a...', 'No es igual a...', 'Mayor que...', etc., with many of them checked. At the bottom of the filter dialog are 'Aceptar' and 'Cancelar' buttons.

### Uso de filtros en Power Query

Fuente: Elaboración propia

En ocasiones la propia estructura de columnas que importamos del origen de datos no se corresponde con nuestras necesidades de forma que tendremos que combinar o dividir las columnas para solucionarlo.

Por ejemplo, si tenemos en una columna los nombres de nuestros clientes y en otra los apellidos, tal vez nos interese unirlos en una única columna. Para ello, seleccionaremos las dos columnas y en la cinta de opciones, en la ficha Transformar, dentro del grupo Columna de texto encontramos la opción Combinar columnas.

RecordId	FirstName	LastName	Fecha de nacimiento	Tienda
1	46440651	Aarón	Álvarez Caletro	null
2	46441119	Abdelilah	-	null   Leganés
3	46845829	Abdelilah	-	null
4	47418636	Abdelilah	Khotabi	06/10/2000   Torrejón
5	46297688	Abdelkerim	El Amrani	21/08/1996
6	47194704	Abdellah	El Meskini	20/01/2002   Leganés
7	46199346	Abdul	Ghemosoune	10/07/1998   Torrejón
8	46400776	Abel	Silva Salazar	null
9	47194751	Abigail	Martínez Rovelo	17/09/2000   Leganés
10	47063066	Abigail	Leon	08/03/1998   Torrejón
11	47072611	Abril	Romero Rolle	20/04/1997
12	46400682	Absa	-	null
13	46718898	achraf	amazian	28/10/1992
14	46368739	Adama	Sackl	04/03/2018
15	46685883	Adil	Tayfi	26/04/1995
16	47194844	Adrián	Arras	10/07/2000   Leganés
17	47194822	Adrián	Beranilla	21/04/2001   Leganés

### Combinar columnas en Power Query

Fuente: Elaboración propia

Cuando los cambios que vamos a hacer en las filas o columnas son muchos, puede ser lo más recomendable referenciar la consulta y hacer los cambios sobre la nueva consulta. Para crear una referencia, solo tenemos que hacer clic con el botón derecho sobre la consulta a referenciar.

Cuando hemos terminado de transformar las consultas, si no queremos cargarlas todas al modelo tabular (por ejemplo, podría haber consultas que hemos utilizado para crear referencias) deberíamos deshabilitarlas haciendo clic con el botón derecho y seleccionando la opción correspondiente del menú contextual. Si tenemos muchas consultas, también podemos agruparlas para que resulten más sencillo trabajar con ellas.

Una vez que hemos finalizado, solo tendremos que hacer clic en Cerrar y aplicar de la ficha Inicio para salir de Power Query y volver a la pantalla principal de Power BI.



### VIDEOTUTORIAL

En este vídeo podemos ver cómo crear visualizaciones de la información en Power BI cuando ya hemos realizado las modificaciones en Power Query:

<https://vimeo.com/user64513894/review/447106536/9ff9638d34>

## 4.6 El lenguaje DAX en Power BI

El lenguaje DAX (Data Analysis Expressions) fue diseñado para crear objetos y consultar modelos tabulares. Estos modelos se encuentran en Power Pivot, integrado en Excel, SQL Server Analysis Services Tabular y Power BI.

DAX fue desarrollado a partir del lenguaje de Excel y contiene todas sus funciones, sin embargo, no son lenguajes exactamente iguales porque DAX trabaja sobre tablas y columnas y no sobre celdas como Excel. Además, cuando las columnas tienen el mismo tipo podemos establecer relaciones entre tablas diferentes.

Los operadores DAX pueden ser de tipo matemáticos, lógicos y de cadenas, operadores de conjunto y operadores de negación. También existen los operadores lógicos AND y OR. Las expresiones DAX nos permitirán consultar y crear columnas calculadas, medidas y tablas calculadas para enriquecer los modelos tabulares.

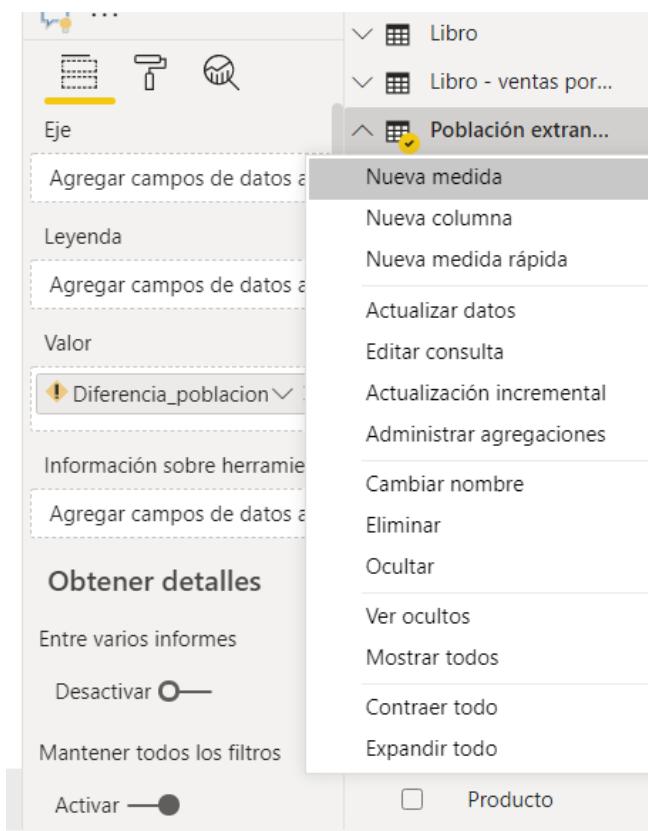


### ENLACE DE INTERÉS

En los siguientes enlaces encontrarás más información sobre el lenguaje DAX de Microsoft

- Guía oficial sobre la sintaxis DAX <https://docs.microsoft.com/es-es/dax/dax-syntax-reference>
- Guía oficial sobre los operadores Dax <https://docs.microsoft.com/es-es/dax/dax-operator-reference>

En DAX hay tres conceptos fundamentales: la sintaxis, las funciones y el contexto. Para activar la barra de fórmulas en la que escribiremos la expresión DAX debemos situarnos en la tabla en la que queremos añadir la nueva columna, activar el menú contextual y seleccionar la opción Nueva medida.



Creación de una nueva medida en Power BI

Fuente: Elaboración propia

También en la ficha Inicio tenemos la opción Nueva medida. Si creamos la medida desde Inicio debemos tener cuidado con seleccionar la tabla en la que queremos incluir dicha medida.

La sintaxis de cada expresión DAX incluye los siguientes elementos:

- El nombre de la columna que crearemos para guardar la información.
- El operador igual.
- La función o la primera columna a la que se hace referencia.
- Los operadores matemáticos necesarios.
- Los demás argumentos necesarios.



Ejemplo de ecuación DAX

Fuente: Elaboración propia

Las funciones en DAX se clasifican de la siguiente manera:

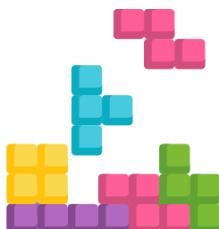
Tipo de función	Descripción
Funciones de inteligencia de tiempo (DAX)	Estas funciones ayudan a crear cálculos que usan el conocimiento integrado sobre calendarios y fechas. El uso de intervalos de fecha y de hora en combinación con agregaciones o cálculos permite crear comparaciones significativas a lo largo de períodos de tiempo comparables relativos a ventas, inventarios, etc.
Funciones de filtro (DAX)	Estas funciones ayudan a devolver tipos de datos específicos, a buscar valores en tablas relacionadas y a filtrar por valores relacionados. Las funciones de búsqueda funcionan usando tablas y relaciones entre ellas. Las funciones de filtro permiten manipular el contexto de datos para crear cálculos dinámicos.
Funciones de información (DAX)	Estas funciones examinan una tabla o una columna proporcionada como argumento en otra función, e indican si el valor coincide con el tipo esperado. Por ejemplo, la función ISERROR devuelve TRUE si el valor al que se hace referencia contiene un error.
Funciones lógicas (DAX)	Estas funciones devuelven información sobre los valores de una expresión. Por ejemplo, la función TRUE permite saber si una expresión que se está evaluando devuelve un valor TRUE.
Funciones matemáticas y trigonométricas (DAX)	Las funciones matemáticas en DAX son similares a las funciones matemáticas y trigonométricas de Excel, si bien existen algunas diferencias en los tipos de datos numéricos que las funciones DAX usan.
Otras funciones (DAX)	Estas funciones realizan acciones únicas que no se pueden definir por medio de ninguna de las categorías a las que la mayoría de las otras funciones pertenecen.
Funciones primarias y secundarias (DAX)	Estas funciones de Expresiones de análisis de datos (DAX) ayudan a los usuarios a administrar los datos que se presentan como una jerarquía de elementos primarios y secundarios en sus modelos de datos.
Funciones estadísticas (DAX)	Estas funciones realizan agregaciones. Además de crear sumas y promedios, o hallar valores mínimos y máximos, DAX también permite filtrar una columna antes de agregar o crear agregaciones basadas en tablas relacionadas.
Funciones de texto (DAX)	Con estas funciones, se puede devolver parte de una cadena, buscar texto dentro de una cadena o concatenar valores de cadena. Existen más funciones para controlar los formatos de fechas, horas y números.



## ENLACE DE INTERÉS

En el siguiente enlace encontrarás el detalle de las más de 200 funciones que podemos usar en DAX

<https://docs.microsoft.com/es-es/dax/dax-function-reference>



## EJEMPLO PRÁCTICO

Supongamos que queremos sumar el total de una columna de una tabla, por ejemplo, las ventas totales de nuestro negocio del año pasado.

Con la función SUM() podremos sumar el total de la columna, pero no nos permitiría agregar ningún filtro, de forma que necesitaremos utilizar otra función DAX.

Para recuperar los datos del año pasado usaremos la función SAMEPERIODLASTYEAR() así que para obtener el resultado que estamos buscando tendremos que utilizar una expresión que contenga tanto SUM() como SAMEPERIODLASTYEAR().

En este caso lo que tendremos que hacer es aplicar la función CALCULATE(). Esta función se expresa de la siguiente forma:

CALCULATE(<expresión a evaluar>;<filtro1>,<filtro2>...)

En este ejemplo, la expresión a evaluar será la suma de las ventas y el filtro el periodo, de forma que la expresión DAX final será:

Venta año anterior = CALCULATE(SUM(Ventas[Importe]);  
SAMEPERIODLASTYEAR(Fechas[Fecha]))

## 5. TENDENCIAS EN BUSINESS INTELLIGENCE

*En el proyecto en el cual estás trabajando, tu director comercial y su equipo de vendedores disponen de listados de contactos e información de clientes que no han sido integrados en ningún sistema corporativo.*

*Piensas que con una plataforma como Zapier puedes recopilar e integrar toda esa información en el CRM corporativo para hacer esa información disponible, buscable y explotable con objeto de mejorar el rendimiento de las ventas.*

Como dice el aforismo, la mejor forma de adivinar el futuro es creándolo. Pero, dado que, de un modo literal, no estamos en disposición de ello, vamos a compartir cuáles consideran como tendencias de futuro en business intelligence tanto los analistas más reputados del mercado, como los fabricantes de software, que si se encuentran en disposición de crear ese futuro.

Si tomamos en cuenta lo que indica el fabricante de software de business intelligence Tableau, que también estudiamos en estas unidades, las principales tendencias del momento en business intelligence son:

- Las herramientas de inteligencia artificial son cada vez más importantes en su aplicación al business intelligence, pero las organizaciones ya no están dispuestas a aceptar modelos de caja negra, sino que exigen que los resultados de la inteligencia artificial sean capaces de explicar y justificar la motivación de sus conclusiones.
- Los usuarios se han acostumbrado al uso del lenguaje natural como medio de interacción con sus ordenadores, teléfonos, vehículos, etc., y, por tanto, también esperan que el lenguaje natural sea el canal de interacción con el que realizar consultas a los sistemas de business intelligence.
- La analítica del business intelligence para ser realmente útil debe ser contextual y situar los datos allí donde se necesitan en cada momento.
- Al igual que se han afianzado en el pasado las colaboraciones, muchas veces denominadas como partnerships, entre el sector público y el privado, por ejemplo, a la hora de desarrollar infraestructuras, también deben fomentarse las colaboraciones para la compartición de datos y la explotación de los mismos.
- Un ejemplo de esto sería el proyecto del Instituto Nacional de Estadística (INE) en colaboración con las operadoras de telefonía

móvil para extraer los datos de ubicación de todos los poseedores de un móvil en España para, tras anonimizarlos, poder llevar a cabo análisis que mejoren las políticas y planes públicos como la vivienda, el transporte, etc.

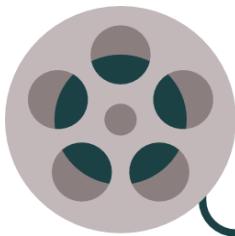
- Los aspectos legales, dado el impacto del Reglamento General de Protección de Datos y las cuestiones éticas, especialmente desde el escándalo de Cambridge Analytica, se van a tener en cuenta más que nunca.
- Debido al crecimiento exponencial de los datos, cada vez es más crítico para las plataformas la administración de estos. Comenzando por la captura y limpieza adecuada de ellos.
- El análisis de los datos debe servir para generar historias. Es decir, el storytelling que, como concepto de marketing y comunicación ha sido ampliamente difundido, llega también a los datos, porque estos deben servir para hablar a las personas: clientes, socios, opinión pública, etc. Entre las herramientas disponibles, las capacidades de visualización son las más importantes para disponer de esa capacidad comunicativa.
- El análisis de datos llegará a todo tipo de organizaciones y se consolidará en aquellas a las cuales ya había llegado.
- El data scientist o, científico de datos, se consolida como la profesión más demanda del sector y una de las mejor pagadas en general en el mercado tecnológico.
- El tránsito de los datos a la nube no solo no se detiene, sino que se acelera.



#### ENLACE DE INTERÉS

Puedes conocer todos los detalles de este informe, así como de los que vayan surgiendo en el futuro de la mano de Tableau, desde aquí:

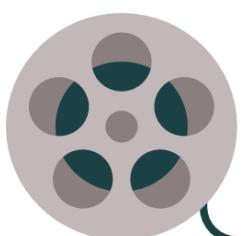
<https://www.tableau.com/es-es/reports/business-intelligence-trends>



### VIDEO DE INTERÉS

En este vídeo encontrarás una noticia sobre el estudio masivo del INE a través de los datos de las operadoras de telecomunicaciones.

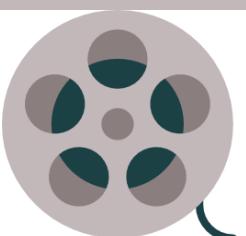
<https://www.youtube.com/watch?v=TDFjY72TxY4>



### VIDEO DE INTERÉS

En este vídeo, varios expertos de Luca (Grupo Telefónica) nos cuentan de primera mano que significa ser data scientist.

[https://www.youtube.com/watch?v=Egnsj\\_iMfaY](https://www.youtube.com/watch?v=Egnsj_iMfaY)



### VIDEO DE INTERÉS

Aprende sobre el storytelling de la mano de una gran experta de LUCA, el área de Telefónica especializada en datos en este vídeo:

<https://www.youtube.com/watch?v=9u5ilJUekdo>

## 5.1 RPA

RPA significa en inglés Robotic Process Automation o, lo que es lo mismo, automatización robótica de procesos. Es un conjunto de herramientas tecnológicas que facilita el que determinados procesos realizados por operadores humanos, normalmente repetitivos y de escaso valor añadido, puedan ser sustituidos mediante un software correctamente parametrizado.

Realmente, la automatización de procesos es uno de los principales objetivos de la informática desde hace décadas, pero el concepto RPA se ha popularizado mucho recientemente para reflejar lo sencillo que resulta

actualmente mediante el uso de diversas plataformas llevar a cabo esta automatización de procesos mediante la aplicación de software.

Es verdad que la palabra robot utilizada en castellano evoca una idea más industrial, de robots físicos y con gran potencia mecánica, utilizados en cadenas de montaje, etc. Se usa la metáfora del robot porque realmente los objetivos perseguidos son los mismos, que es alcanzar una mayor productividad sin reducir la calidad. Son especialmente utilizados para poder llevar a cabo proyectos de transformación digital.

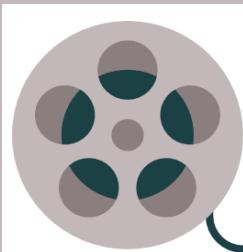
La automatización de los procesos genera una gran cantidad de datos que pueden ser analizados, por lo que RPA y business intelligence están muy vinculados.

Además, RPA también se puede utilizar para mejorar los repositorios de datos en la organización, ya que puede facilitar la extracción e integración de datos con origen en distintas fuentes, por lo que la calidad del análisis realizado sobre ellos será mejor.

La RPA tienen dos funciones interesantes desde el punto de vista de los datos:

- Ayuda a crear metadatos: A medida que los robots completan las tareas para las cuales han sido programados, registran su progreso, así como los problemas que han detectado para realizar el trabajo encomendado. Estos datos pueden ser utilizados tanto por el cliente, como por el proveedor de RPA, para identificar los fallos y mejorar el rendimiento de los bots.
- Habilitan el acceso a los datos de los sistemas heredados: Dado que las tareas que interactúan con los sistemas de información antiguos suelen ser lentas, los RPA pueden encargarse de la recopilación de los datos en ellos almacenados.

Por tanto, RPA no solo mejora las capacidades analíticas, sino que ayudan a la recopilación de datos. Incluso, muchos proveedores de RPA destacan este aspecto recalmando que uno de sus principales beneficios es el mecanismo de federación de datos que agregan



### VIDEO DE INTERÉS

En este vídeo podrás saber qué es RPA y, además, ver un ejemplo de uso de una plataforma RPA

<https://www.youtube.com/watch?v=G6VtYd7brd8>

## 5.2 Zapier

Actualmente realizamos muchas tareas repetitivas que nos consumen tiempo y que son fácilmente automatizables. Casi todas las herramientas que se encuentran en la web cuentan con una API que facilita el intercambio de información con otras herramientas. Zapier es una aplicación que permite conectar las API de más de 2.000 herramientas web diferentes.

**Connect your apps and automate workflows**

Easy automation for busy people. Zapier moves info between your web apps automatically, so you can focus on your most important work.

Sign Up With Google  
Sign up with Facebook  
OR  
Work Email  
First Name Last Name  
Get Started Free

By signing up, you agree to Zapier's [Terms of Service](#).

Home de acceso a la plataforma Zapier

Fuente: [zapier.com](http://zapier.com)

El origen de Zapier se encuentra en 2011, cuando sus fundadores, Wade Foster, Bryan Helmig y Mike Knoop participaron en el primer Startup Weekend Columbia. Animados por los comentarios que recibieron, los tres decidieron seguir adelante con su idea. Durante los siguientes meses, los tres trataron de sacar adelante el prototipo inicial y encontrar usuarios beta. Al principio, los usuarios no podían hacer la integración de las aplicaciones por su cuenta, sino que los propios desarrolladores llamaban a los potenciales usuarios, les preguntaban qué querían integrar y ellos -los propios fundadores- eran quienes hacían que esa integración funcionase.

También, participaban activamente en los foros de las principales herramientas para conocer con qué otras aplicaciones querían enlazarlas. De esta forma, a través de los foros de usuarios de Dropbox, descubrían que a muchos les gustaría poder guardar en una carpeta concreta los archivos añadidos a una nota de Evernote o los recibidos en determinados emails. A pesar de que en sus primeras versiones Zapier era un producto inacabado y todavía se producían muchos errores, gracias a que se centraban en adaptar cada integración a cada uno de sus clientes, consiguieron fidelizarlos y al mismo tiempo conseguían la información necesaria para mejorar Zapier.

Un año después del fin de semana en el que surgió, en 2012, Zapier era una beta funcional en la que mil personas habían pagado para tener acceso, y otras 10.000 personas estaban en lista de espera para probarla y contaban con 25 integraciones. Dos años más tarde, ya tenía 600.000 usuarios y en 2019 contaba con más de 3 millones.

Las integraciones entre aplicaciones web existen desde hace mucho, sin embargo, durante mucho tiempo era necesario tener amplios conocimientos sobre el funcionamiento de cada una de las aplicaciones para poder realizar esas integraciones. Zapier facilita que esas integraciones puedan ser realizadas por casi cualquier persona.

Actualmente, Zapier ayuda a conectar más de 2.000 aplicaciones web y redes sociales a través de sus respectivas API.



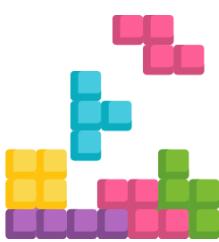
**VIDEOTUTORIAL**

Con este vídeo podrás ver cómo registrarte en Zapier:

<https://vimeo.com/user64513894/review/447107536/a4b22ebff9>

Las API son parte muy importante de internet tal como lo conocemos hoy en día, porque sirven para poder comunicar diferentes aplicaciones unas con otras. De este modo, por ejemplo, podemos tomar información de un mail de Gmail y utilizar esa información para completar una tabla de Google Sheets.

Zapier agrega constantemente nuevas aplicaciones y redes sociales a su plataforma. Por este motivo, Zapier se convirtió en una de las herramientas principales para comunicar diferentes API sin necesidad de saber programar.



### EJEMPLO PRÁCTICO

Si recibimos un contacto a través de un formulario en nuestro sitio web y el formulario está desarrollado con una tecnología que utiliza una API integrada con Zapier, podemos hacer que la información de ese formulario se complete automáticamente en una hoja de cálculo de Google Sheets o se añada a una lista de contactos de Mailchimp.



### VIDEOTUTORIAL

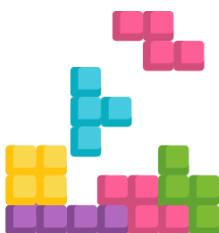
En este vídeo te explicamos el entorno de trabajo de Zapier:

<https://vimeo.com/user64513894/review/447108898/a4f4dd0e9f>

Zapier funciona en base a Zaps. Un Zap es básicamente una tarea específica que deseamos realizar.

De esta forma, los Zaps utilizan las API de diferentes aplicaciones para integrarlas. Todos los Zaps funcionan en base a disparadores y acciones: un disparador es un evento que debe ocurrir en una aplicación para iniciar la tarea del Zap; una acción, como su nombre lo indica, es la acción que deseamos realizar con Zapier de forma automática una vez que ocurra el disparador.

Los disparadores y acciones son la base de Zapier. Cada Zap creado debe tener, al menos, un disparador y una acción. Los disparadores son eventos que activan nuestro Zap. Por ejemplo, recibir un correo electrónico de una dirección determinada puede ser un disparador. Otro disparador podría ser, por ejemplo, enviar una campaña de e-mail Marketing a través de MailChimp. O, por ejemplo, suscribir un nuevo usuario a una lista específica de MailChimp. Cada Zap debe poseer un disparador, es decir, cuál es el evento que disparará otra tarea relacionada.



### EJEMPLO PRÁCTICO

Siguiendo con el ejemplo de MailChimp, podemos hacer que cada vez que se suscribe un nuevo usuario a nuestra lista de MailChimp, también se ingresen los datos del nuevo usuario suscrito en una tabla de Google Sheets.

La suscripción sería el disparador y completar los datos en Google Sheets sería la acción.



### VIDEOTUTORIAL

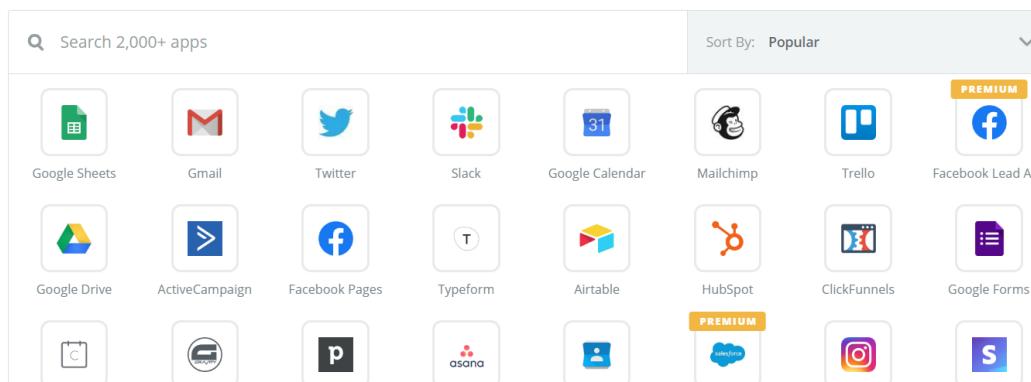
En este vídeo podemos ver un ejemplo de cómo crear un Zap con el que se van a crear contactos de Google cuando recibimos un formulario y cómo guardar también esos datos en una hoja de cálculo.

<https://vimeo.com/user64513894/review/447110212/9f5b88ff3d>

Cada Zap debe poseer, al menos, una acción, pero Zapier también brinda la posibilidad de crear Zaps de múltiples pasos, es decir, que a partir de un disparador ocurran múltiples acciones de automatización. Utilizando la lógica de disparadores y acciones podemos lograr automatizar las tareas que realizamos de forma cotidiana y de ese modo ahorrar tiempo y recursos.

Cuando vamos a la pestaña Explore podemos ver las aplicaciones más populares con las que podemos empezar a trabajar. Podemos ordenar las aplicaciones en orden alfabético, las añadidas de forma más reciente, o las aplicaciones premium, así como también organizarlas por categorías, como contabilidad, publicidad o analítica.

### Pick Apps to Explore Workflow Ideas



The screenshot shows a grid of application icons categorized under 'Popular'. The icons include Google Sheets, Gmail, Twitter, Slack, Google Calendar, Mailchimp, Trello, Facebook Lead Ads, Google Drive, ActiveCampaign, Facebook Pages, Typeform, Airtable, HubSpot, ClickFunnels, Google Forms, Asana, LinkedIn, and Instagram. A 'PREMIUM' badge is visible next to several icons.

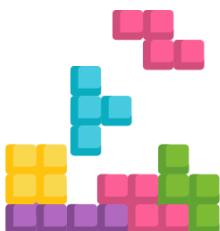
#### ❤️ Popular Zaps on Zapier

Aplicaciones disponibles para su integración en Zapier

Fuente: Zapier.com

Por ejemplo, cómo podemos ver, Google Sheets es una aplicación muy popular para automatizar tareas. Esto es muy útil, porque podemos tanto añadir información, como modificarla o extraerla para comunicar varias aplicaciones entre sí.

Dicho de otra manera, podemos extraer los listados de correos electrónicos de una hoja de Google Sheets e importar dichos correos electrónicos como suscriptores en una lista de MailChimp. Podemos trabajar con nuestras páginas de Facebook o Slack o, por ejemplo, Google Calendar para agendar reuniones de forma automática. También podemos trabajar con redes sociales, por ejemplo, Twitter o LinkedIn, para hacer publicaciones de contenidos de forma automática.



### EJEMPLO PRÁCTICO

Podemos hacer también automatizaciones interesantes con Paypal. Supongamos que tenemos una hoja de Google Sheets con una columna con el estado de un cliente.

Podemos hacer que el estado cambie de activo a inactivo si es que el cliente no nos realiza un pago y luego podemos automatizar una tarea de enviar una solicitud de pago a los clientes cuyo estado haya cambiado de activo a inactivo.

Una vez que el pago se ha recibido de forma correcta, volver a activarlo de forma automática en la hoja de cálculo. De este modo, podremos administrar nuestros cobros de forma automática a través de un documento de Google Sheets y la aplicación premium de PayPal de Zapier.

Zapier ofrece muchas opciones para automatizar nuestro negocio y es probable que con un poco de creatividad podamos encontrar varias tareas que actualmente hacemos a mano y poder realizarlas de una forma mucho más eficiente y sin consumir tiempo de trabajo.



### VIDEOTUTORIAL

En este vídeo te explicamos cómo crear zaps con filtros para aumentar la precisión de la automatización de las tareas.

<https://vimeo.com/user64513894/review/447112020/5c64e>

[cd992](#)

## RESUMEN FINAL

A lo largo de esta unidad hemos realizado un recorrido por la sociedad y la tecnología de las últimas décadas hasta llegar al presente. Incluso, hemos tenido ocasión de realizar un ejercicio de prospectiva y anticipar que nos depararán los próximos años en materia de gestión de datos, analítica e inteligencia de negocio en general.

Comenzamos nuestro recorrido analizando la evolución social que ha llevado a que los datos y la información se hayan convertido en lo que para muchos es el petróleo del siglo XXI.

Continuamos la presente unidad identificando cómo los sistemas y tecnologías de la información han evolucionado e innovado para aportar más valor a todo tipo de organizaciones, especialmente a las empresariales, en lo que tiene que ver con la captura de datos y su análisis posterior.

Hemos profundizado en las distintas tecnologías que son necesarias para desplegar un sistema de business intelligence, con especial énfasis en el data warehouse, por ser este la piedra angular de las plataformas de inteligencia de negocio.

Pero no nos hemos ocupado solamente del data warehouse y todas sus variantes, sino que hemos analizado como otros elementos son necesarios para aportar un gran valor a los distintos estamentos de la empresa, sean usuario de alto nivel directivo u otros de perfil más operativo: DSS, sistemas de reporting, EIS, etc.

Finalmente, hemos contextualizado el valor que tienen los algoritmos de clustering para mejorar los análisis de negocio que una empresa puede realizar y también hemos recordado los ya conocidos algoritmos de clasificación y cómo se relacionan con la analítica de negocio.

En paralelo a los aspectos de business analytics, hemos tenido ocasión de introducirnos y aprender a usar una conocida herramienta de mercado que es Power BI. Esta plataforma de análisis ha ganado una importante cuota de mercado. No en vano, es un producto de la todopoderosa Microsoft, que lo está potenciando al máximo en el mercado empresarial.

Finalmente hemos cerrado esta unidad con una mirada al futuro que nos depara esta apasionante actividad que es la inteligencia de negocio.