

Gobierno del Dato y Toma de Decisiones

Tema 2. Business intelligence y datos maestros

Índice

Esquema

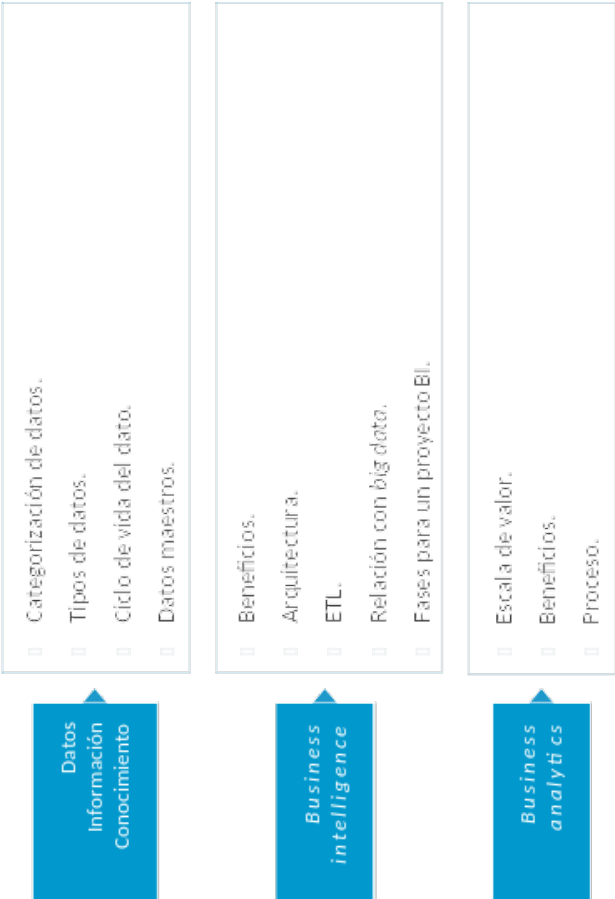
Ideas clave

- 2.1. Introducción y objetivos
- 2.2. Datos, información y conocimiento
- 2.3. Datos maestros
- 2.4. Inteligencia de negocios
- 2.5. Business intelligence vs. business analytics
- 2.6. Referencias bibliográficas

A fondo

- Introducción al business intelligence
- Business intelligence y business analytics
- Curso introductorio de business intelligence y business analytics
- PowerData

Test



2.1. Introducción y objetivos

¿Es lo mismo datos, información o conocimiento? La tendencia inicial es a responder afirmativamente a este interrogante, pero la respuesta correcta es «No». En este tema veremos en qué radica la diferencia principal y cómo los datos son la base de una buena generación de conocimiento. Veremos que existen diferentes tipos de datos, que el dato tiene un ciclo que debería cumplirse para poder garantizar seguridad y calidad. Además, veremos la importancia de crear repositorios de datos maestros, las ventajas de esta práctica y algunos pasos a seguir para su implementación.

Luego, seguiremos con las generalidades de los datos y entraremos en el mundo del *business intelligence* y el *business analytics* viendo sus principales diferencias y algunos pasos que nos faciliten su implementación.

Los objetivos de este tema son:

- ▶ Diferenciar los conceptos de dato, información y conocimiento.
- ▶ Analizar los pasos a seguir para implementar un repositorio de datos maestros.
- ▶ Comprender la importancia del ciclo que debe cumplir el dato en una estrategia de gobierno de datos.
- ▶ Conocer las semejanzas y diferencias entre *business intelligence* y *business analytics*.

2.2. Datos, información y conocimiento



Figura 1. Preguntas sobre datos e información.

¿Es lo mismo datos, información o conocimiento? La tendencia inicial es a responder afirmativamente a este interrogante, pero la respuesta correcta es «No».

Si empezamos a indagar un poco sobre ellas, las tres tienen significado diferente:

- ▶ **Los datos** son elementos sin procesar, sacados de la realidad que a su vez generan nuevos elementos y que por sí solos no generan nuevo conocimiento. Ejemplos de datos: el precio de un producto, la edad, el nombre de una persona, etc.
- ▶ **La información** es el principio del conocimiento. Son datos con un significado o función especial o el resultado de combinar diferentes datos, es decir, son datos con contexto.
- ▶ **El conocimiento** es la información analizada que hace nuevos aportes a un área específica.



Figura 2. Cómo se integran los datos y la información. Fuente: Weller, 2010.

Los datos son la principal fuente de información para el análisis de grandes volúmenes, por lo que es fundamental que las empresas realicen una selección lo más adecuada y cuidadosamente posible. La categorización de los datos es importante para cualquier proyecto, en especial cuando se trabaja con grandes volúmenes (*big data*). La principal categorización de los datos se basa en su estructura, en la que nos encontramos principalmente dos posibilidades:

- ▶ **Estructurados:** estos datos son aquellos que tienen una estructura definida y que no cambian independientemente de cuál sea su origen. Es decir, son aquellos datos que poseen un modelo (o estructura) definido. Entre estos tipos de datos podemos encontrar registros de bases de datos, que son el ejemplo más típico, datos de sensores o los que se obtienen a partir del API de Twitter.
- ▶ **No estructurados:** aquellos datos que no disponen de una estructura bien definida. Es decir, no poseen un modelo (o estructura) definido o no están ordenados de ninguna forma. Entre estos tipos de datos podemos encontrar fotografías, vídeos o documentos de texto (Word, PDF, etc.).



Figura 3. Introducción a la categorización de datos. Fuente: elaboración propia.

En cuanto a cómo se han generado los datos, nos podemos encontrar con los siguientes tipos de datos:

- ▶ **Creados:** son aquellos generados por la propia empresa a través de los sistemas de información.
- ▶ **Compilados:** son aquellos que se utilizan de otras grandes bases de datos, como censos electorales, información obtenida de las administraciones públicas en salud, vivienda, impuestos, etc.
- ▶ **Experimentales:** son los generados por simulaciones o pruebas para determinar la validez de los sistemas.

En la Figura 4 se muestran algunas fuentes de información que hoy en día se han ampliado gracias a las nuevas tecnologías de la información.

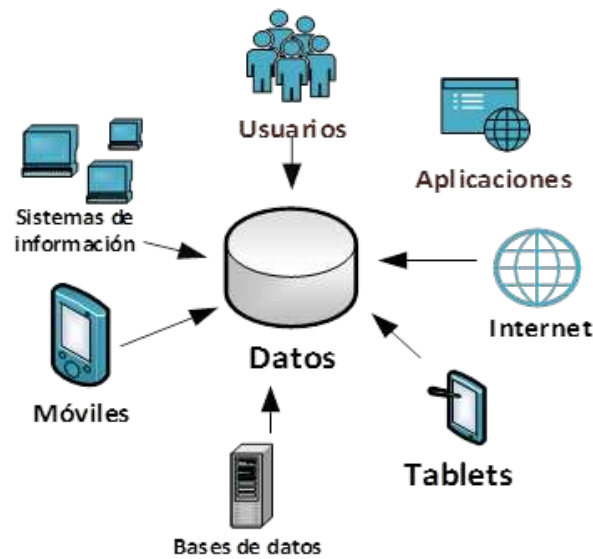


Figura 4. Fuente de datos.

- ▶ **Web (Internet) y medios sociales:** son aquellos que tienen origen en la red. Según los expertos, es la fuente más grande del *big data* y una de las más utilizadas en la actualidad. Se genera información en los clics de los vínculos y elementos, así como en las búsquedas que se hacen, las publicaciones en las redes sociales (Twitter, Facebook, LinkedIn...) y el contenido web como páginas, enlaces o imágenes.
- ▶ **Tabletas y móviles (*smartphone*):** son dispositivos móviles que permiten acceder a información desde cualquier parte gracias a su tamaño, las aplicaciones y su acceso a Internet. Entre todas las fuentes de datos utilizables, el teléfono inteligente o tableta es el que más potencial tiene.

Gracias a estos dispositivos, compañías como Google y Apple pueden saber dónde se encuentra una persona en cada momento y, de esta manera, poder conocer los gustos de comida, vestuario y diversión, entre otros.

- ▶ **Bases de datos y sistemas de información:** una base de datos es un almacén en el que se pueden organizar datos, para evitar la redundancia y mejorar el acceso a estos. Diferentes programas y usuarios deben poder utilizar los datos guardados. De ahí el término base «sistema de información».

Tipos de datos

Para los diferentes procedimientos que se realizan en la inteligencia de negocios, es importante definir los tipos de datos que se van a usar para obtener conocimiento.

En la empresa se pueden encontrar los siguientes:

- ▶ **Numéricos:** aquellos valores enteros o reales. Por ejemplo: la edad, el salario y las horas trabajadas, entre otros.
- ▶ **Fecha/hora:** se identifican como campos de fecha/hora aquellos valores cuyo contenido encaja con formatos de fecha/horas más usuales. Sus componentes pueden ser: año, mes, día, hora, minutos o segundos.
- ▶ **Texto:** sus valores tienen texto libre, es decir, que no está limitado a un subconjunto de etiquetas. Por ejemplo, el nombre de los empleados y el nombre del departamento, entre otros.
- ▶ **Alfanuméricos:** son una combinación de los datos numéricos y datos de tipo texto como, por ejemplo, el número de DNI, la dirección, etc.
- ▶ **Booleano:** es un tipo con solo dos valores: verdadero y falso (Gagliardi et al., 1999).

La mayoría de las veces los datos no pueden ser utilizados de la manera en que se encuentran almacenados, pues pueden encontrarse en las ya mencionadas bases de datos o en archivos planos, entre otros. Debido a la poca rigidez de los sistemas de calidad y control al introducirlos a los diferentes sistemas, existen diferentes tipos de errores: datos incompletos, datos con ruido, datos inconsistentes o duplicados.

Es necesario tener un proceso de **calidad** de los datos para minimizar los errores y poder obtener un mayor rendimiento en el análisis; de lo contrario, los algoritmos generalmente ignoran el dato y se pierde información valiosa.

El autor Dorian Pyle define la preparación de los datos como «la manipulación y

transformación de los mismos sin refinar para que la información contenida en el conjunto de datos pueda ser descubierta o estar accesible de forma más fácil» (Pyle, 1999).

Los componentes para la preparación de datos son: limpieza, integración, transformación y reducción.

- ▶ **Limpieza:** esta tarea puede involucrar: completar los datos faltantes, resolver los problemas de ruido y eliminar valores extremos o corregir los datos inconsistentes (Han et al., 2006).
- ▶ **Integración:** los datos pueden provenir de diferentes fuentes y, por esta razón, es necesario agruparlos en diferentes tablas para crear información homogénea. Los datos que provienen de diferentes fuentes pueden tener diferentes formatos y escalas. La recopilación tiene que ser coherente. Frecuentemente, esta integración de datos se realiza en una base de datos (Herrera et al., 2004).
- ▶ **Transformación:** en este paso se crean nuevos atributos a partir de los atributos originales. Esta transformación puede facilitar una mejor interpretación de la información (Lin, 2002). Por ejemplo, el índice de masa corporal en datos médicos se calcula con el peso y la altura de una persona. Otros ejemplos de transformación de datos son la discretización, la normalización y derivación.
- ▶ **Reducción:** la reducción de la dimensionalidad consiste en aplicar una transformación para conseguir una representación reducida de los datos originales sin perder información. En esta parte se deben escoger aquellas variables o atributos que influyan en la obtención de conocimiento.

La capacidad de la organización de gestionar los datos de forma eficiente depende de que los datos cumplan un ciclo de vida y se garanticen las mejores prácticas en cuanto a calidad y seguridad en cada una de ellas. El ciclo de vida se puede ver en la Figura 5.

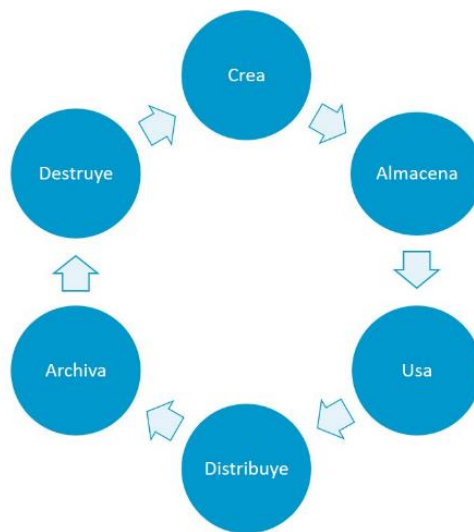


Figura 5. Ciclo de vida del dato en el gobierno de datos. Fuente: adaptado de Dataworks, s.f.

2.3. Datos maestros

Uno de los mayores problemas a la hora de consolidar los datos es que estos se encuentran en diferentes fuentes y pueden llegar a tener valores diferentes. Por ejemplo, un mismo cliente puede estar en dos bases de datos y tener una dirección o un teléfono diferente en cada una de ellas. La gestión de datos maestros (MDM, por sus siglas en inglés de *master data management*) nace como solución a este problema. El objetivo es que los datos críticos se encuentren en un solo sitio, simplificando de esta forma el intercambio de datos entre personas y departamentos de una misma organización (Trejo, 2020).

Los datos maestros se crean en general para unificar datos de clientes, empleados, proveedores y productos. Contienen la información central compartida dentro de la organización, la administración de los datos maestros es clave y necesaria, debe ejecutarse en coordinación con el grupo de gobierno de datos.

¿Es importante utilizar MDM?

Un error en un dato puede causar errores en todas las decisiones de una empresa. Si los datos maestros se crean pero no se administran, esto puede repercutir en todas las aplicaciones que utilizan dichos datos, así que es muy importante crear datos maestros y administrarlos. Un error muy típico se produce cuando un cliente cambia su dirección y esta es actualizada en una fuente de datos, pero no en todas, y nunca recibe notificaciones. Tiempo después, se dan cuenta de que la dirección no ha sido modificada en todas las fuentes de datos. Esto puede causar impagos o problemas serios al cliente. **Los datos maestros deben ser correctos y consistentes.**

Combinar datos maestros puede llegar a ser una tarea compleja. Por ejemplo, si se fusionan dos empresas, cada una de ellas tiene una estructura diferente en sus bases de datos. El mismo cliente puede tener varios nombres, números de

identificación diferentes, direcciones, teléfonos... Es necesario utilizar herramientas existentes que ayuden en esta tarea. Consolidar el maestro de datos es un desafío, pero a la vez tiene muchos beneficios: datos de factura consolidados, *marketing* más efectivo y codificación única de artículos. La Figura 6 explica los pasos más importantes a seguir para la creación de un repositorio de datos maestros.

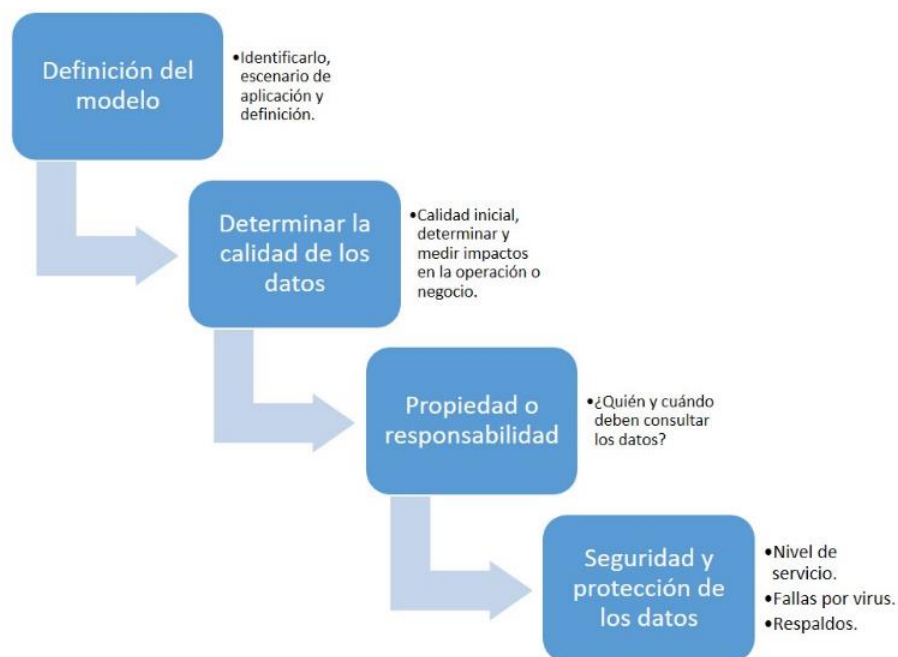


Figura 6. Vista *top down* de los pasos a seguir para un maestro de datos. Fuente: adaptado de Trejo, 2020.

La eficacia de la gestión de datos maestros se puede medir mediante varios criterios:

- ▶ **Disponibilidad de datos:** medido del 0 al 1, de acuerdo con lo interesantes que pueden ser los datos generados para cierta consulta.
- ▶ **Calidad de los datos:** rapidez, completitud y veracidad. También con una medición de 0 a 1.
- ▶ **Desempeño de la consulta de datos:** datos en tiempo real, incluyendo datos agregados. Medición entre 0 y 1.

2.4. Inteligencia de negocios

Si se cuenta con un repositorio de datos maestros unificados, es más fácil implementar una estrategia de inteligencia de negocios. Hasta hace algunos años estos datos maestros se construían en el proceso BI, ahora con el gobierno de datos la idea es crearlos antes de cualquier proceso BI, BA u otro que implique generación de conocimiento.

El *business intelligence* (BI), que se traduce como «inteligencia de negocios», es un **proceso de intercambio** para explorar y analizar información estructurada de la empresa o sobre una determinada área (con frecuencia almacenada en un *data warehouse*), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conocimiento para el mejoramiento de la empresa (Teixeira et al., 2019).

El proceso del *business intelligence* incluye la comunicación de los descubrimientos y la ejecución de los cambios. Las áreas que abarcan, por lo general, son clientes, proveedores, productos, servicios y competidores (véase en el subapartado de Webgrafía el ejemplo de Gather).

Implementar soluciones de inteligencia de negocios dentro de la empresa ayuda en las decisiones que se toman. También contribuyen a **nivel interno**, para apoyar la gestión del personal (Sharma et al., 2009), y a nivel externo, producir ventajas sobre sus competidores (Valenzuela, 2007).

En algunas ocasiones no se pueden lograr todos los beneficios que tiene el *business intelligence* debido al proceso que se lleva a cabo al implementar un proyecto de estas características. Se pueden cometer errores en la definición del planteamiento de las necesidades de conocimiento de la empresa y, al no determinar bien los problemas de información que necesitan solución, generalmente causan el fracaso

del proyecto.

¿Quién necesita el
business intelligence
?

Todas aquellas personas de la empresa que tienen que tomar decisiones. Dependiendo del tipo de negocio, se deben hacer las preguntas necesarias para responder y establecer el modelo de *business intelligence* que mejor se adapte.

Beneficios del
business intelligence

Uno de los objetivos básicos de los sistemas de información es contribuir a la toma de decisiones. Cuando se requiere tomar una decisión, es necesario pedir o buscar información, que servirá para minimizar la incertidumbre. En todo caso, no todos los responsables recogen la misma información: depende de factores como la experiencia, la formación y la disponibilidad, entre otros.

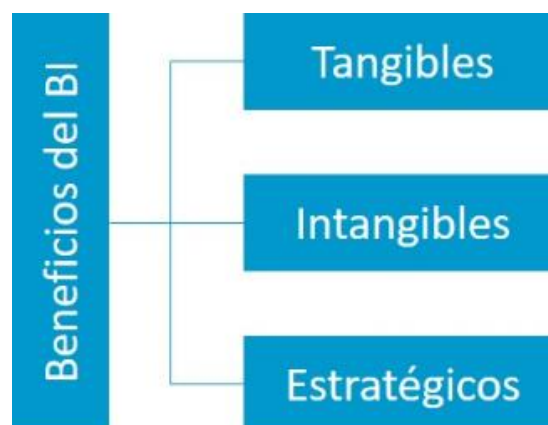


Figura 7. Beneficios del *business intelligence*. Fuente: elaboración propia.

Del mismo modo, los responsables pueden necesitar recoger más o menos información dependiendo del tipo de problema por resolver. A partir de los datos que proporciona el sistema de *business intelligence*, se puede descubrir nuevos aportes (conocimiento). Los beneficios pueden ser de distintos tipos (Figura 7) (Puklavec et al., 2018):

- ▶ **Beneficios tangibles:** son aquellos que la empresa puede cuantificar y que le aportan beneficios económicos. Ejemplo: reducción de costes de producción, generación de nuevos ingresos, reducción en tiempo de producción, evitar pérdidas de clientes o materia prima, aumentar la rentabilidad.
- ▶ **Beneficios intangibles:** son aquellos que no se pueden cuantificar, pero que aportan valor agregado a los servicios o productos y mejoran la posición competitiva. Ejemplo: mejorar la atención al cliente, aumentar la satisfacción del cliente interno y externo, tener información más actualizada.
- ▶ **Beneficios estratégicos:** son aquellos que facilitan la creación de nuevas estrategias, respecto a qué clientes, mercados o con qué productos encaminar los esfuerzos de la empresa. Ejemplo: mejorar la toma de decisiones, identificar clientes potenciales, etc.

Arquitectura del *business intelligence*

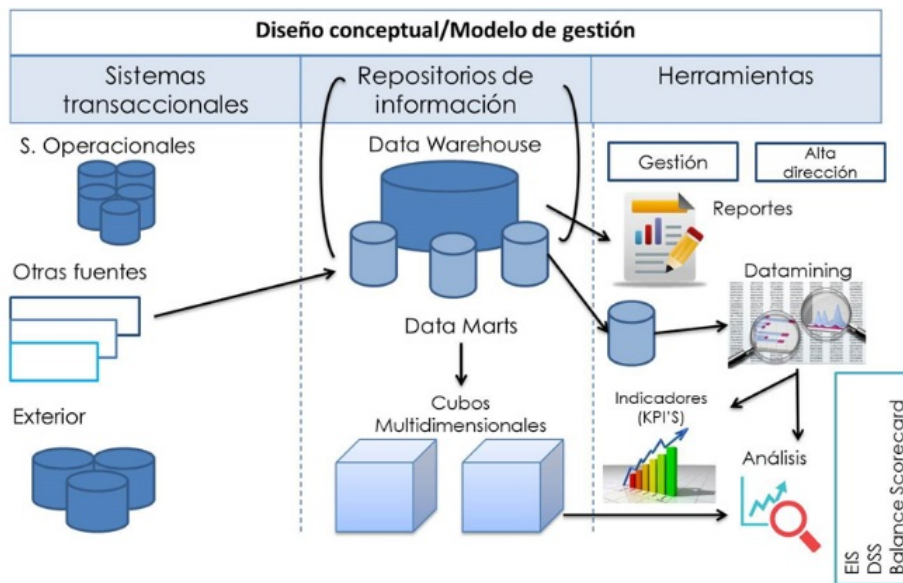


Figura 8. Arquitectura del *business intelligence*. Fuente: IDS2015, 2015.

Una solución de este sistema parte de varias fuentes de datos que suelen ser transformadas estructuralmente para optimizar el análisis, proceso al que se le denomina ETL. Una vez está unificada, la información se almacena en un *data warehouse* que puede servir como base a distintos *data marts*. Los datos almacenados en el *data warehouse* o *data mart* se explotan utilizando herramientas de visualización o *reporting*. Esto lo podemos observar en la Figura 8.

- ▶ **Procesos ETL:** consisten en la extracción, transformación y carga de los datos en el *data warehouse*. Antes de guardarlos ahí, deben ser transformados, limpiados, filtrados y redefinidos. Como se mencionó anteriormente, la información que tienen las empresas en los sistemas transaccionales no está preparada para la toma de decisiones.
- ▶ **Data warehouse:** también llamado almacén de datos, con el *metadata* o diccionario de datos. Se busca almacenar los datos de una forma que facilite y maximice su flexibilidad, facilidad de acceso y administración. Surge como respuesta a las necesidades de los usuarios que necesitan información consistente, integrada, histórica y preparada para ser analizada y apoyar la toma de decisiones.

- **Herramientas OLAP:** para proveer la capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de escenarios en grandes volúmenes de datos. En la actualidad, existen otras alternativas tecnológicas al OLAP. Siguiendo el modelo se deben analizar las tecnologías que permitirán tratar y visualizar la información que reside en un *data warehouse*. En este apartado también se tratarán las herramientas de visualización, ya que en muchas ocasiones van ligadas.

Relación entre *business intelligence* y *big data* (BD)

El *business intelligence* se refiere a las habilidades, tecnologías, aplicaciones y prácticas para la exploración iterativa continua del pasado empresarial para proporcionar información útil en el presente. El *business intelligence* se centra en el desarrollo de nuevos conocimientos y la comprensión del rendimiento empresarial basados en métodos estadísticos.

El término *big data* se usa para caracterizar conjuntos de datos grandes, diversos y que cambian rápidamente, lo que es cada vez más frecuente en todas las organizaciones. *Big data* requiere de sistemas de administración de bases de datos con capacidades más allá de las que se ven en los sistemas estándar basados en SQL.

La mayoría de las definiciones intuitivas de *big data* se centran en el volumen de datos que se producen, a menudo medidos en términos de **tera** (10¹²), **peta** (10¹⁵) o **exa** (10¹⁸) *bytes*. Algunos afirman que se está ingresando en la llamada «Era *Petabyte*» (Anderson, 2008) mientras que otros prefieren hablar de cuántos *exabytes*

de datos se producen cada día (McAfee y Brynjolfsson, 2012).

Sin embargo, aunque el volumen es, sin duda, un aspecto del *big data* (probablemente el menos problemático), a medida que la tecnología se desarrolla, lo que fue grande en el pasado será normal mañana y probablemente se piense que es bastante pequeño en el futuro. Por consiguiente, para comprender qué hace que el *big data* sea diferente, también se debe considerar las dimensiones de la **velocidad**, la **veracidad** y la **variedad**.

- ▶ **Velocidad:** mientras que el volumen se refiere a lo que podría considerarse una reserva de datos, la velocidad se refiere a la que esa acción cambia; por ejemplo, la velocidad a la que se generan los datos, la frecuencia a la que se actualiza o la velocidad a la que son entregados.

Entre los ejemplos de datos de alta velocidad se incluyen datos financieros de mercados bursátiles, datos en tiempo real de sensores y cámaras de vídeo y datos de *stream* generados por visitantes a tiendas en línea.

- ▶ **Veracidad:** la veracidad hace referencia a la incertidumbre de los datos, es decir, al grado de fiabilidad de la información recibida. Para solventar el problema de la veracidad, es necesario conseguir datos de calidad, aplicando soluciones y métodos que puedan eliminar datos imprevisibles que puedan surgir como datos económicos, comportamientos de los consumidores que puedan influir en las decisiones de compra.
- ▶ **Variedad:** aunque tal vez no sea tan obvio como el volumen o la velocidad, en muchos sentidos la variedad plantea el mayor problema para el análisis de *big data*. La variedad se refiere a la cantidad de diferentes fuentes de las que pueden proceder los datos y los formatos, estructuras y semántica que están asociadas a ellas (la estructura se refiere tanto al formato en el que se almacenan los datos como el número y la longitud de los campos, y, más crucial, la semántica que debe asociarse con esos campos). Para que una computadora pueda procesar datos de

manera que sean válidos y significativos para los seres humanos, primero los datos deben codificarse, es decir, un valor semántico —efectivamente un significado— tiene que asignarse a cada elemento de datos (Kimble, 2013).

Si bien estas cuatro variables (las 4 uves) son las más prominentes, revistas especializadas hablan de 5, 8 y hasta 10 uves. En un artículo de febrero de 2017, el portal *Transforming Data with Intelligence* proponía la siguiente lista de uves: volumen, velocidad, variedad, variabilidad, veracidad, validez, vulnerabilidad, volatilidad, visualización y valor (Firincan, 2017).

Las diferencias entre **business intelligence** y **big data** tienen consecuencias sobre cómo están organizadas. Tradicionalmente, los equipos de *business intelligence* están ubicados en organizaciones de consultoría internas, centros de excelencia o departamentos de TI, donde proporcionan a los gerentes y ejecutivos reportes con información bien definida, estable y útil (Burton et al., 2006; Davenport et al., 2012).

Sin embargo, dado que la mayoría de las iniciativas de *big data* carecen de preguntas predefinidas y son de naturaleza mucho más experimental (Casey et al., 2013), los especialistas de *big data* deben organizarse para que estén cerca de los productos y procesos en las organizaciones, es decir, que compartan negocios y unidades (Davenport et al., 2012). A continuación, se muestran algunas semejanzas y diferencias en las áreas de competencia de *business intelligence* y *big data*.

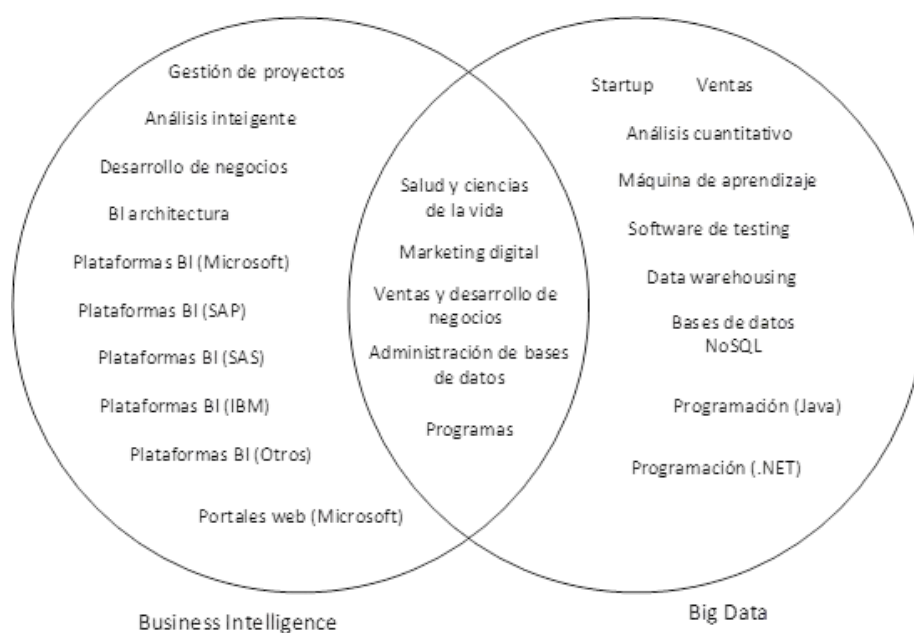


Figura 9. Semejanzas y diferencias en las áreas de competencia de *business intelligence* y *big data*.

Fuente: Debortoli et al., 2014.

¿Cómo construir un ecosistema BI?

En la Figura 10 podemos ver las diferentes etapas que se deberían seguir para poder implementar un proyecto BI en una empresa.

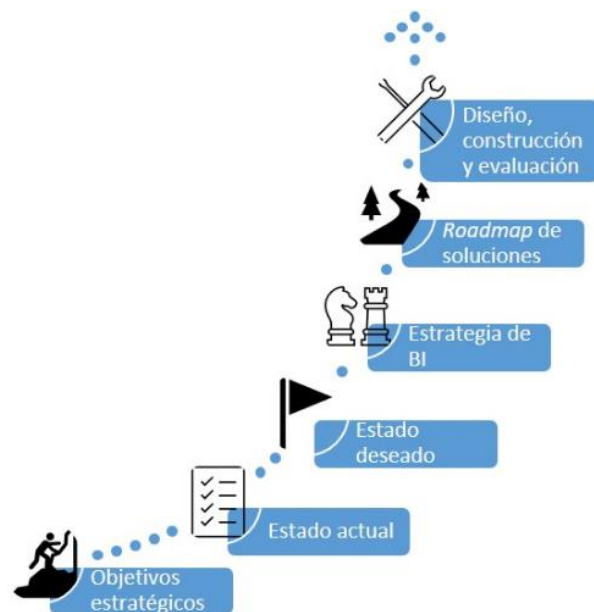
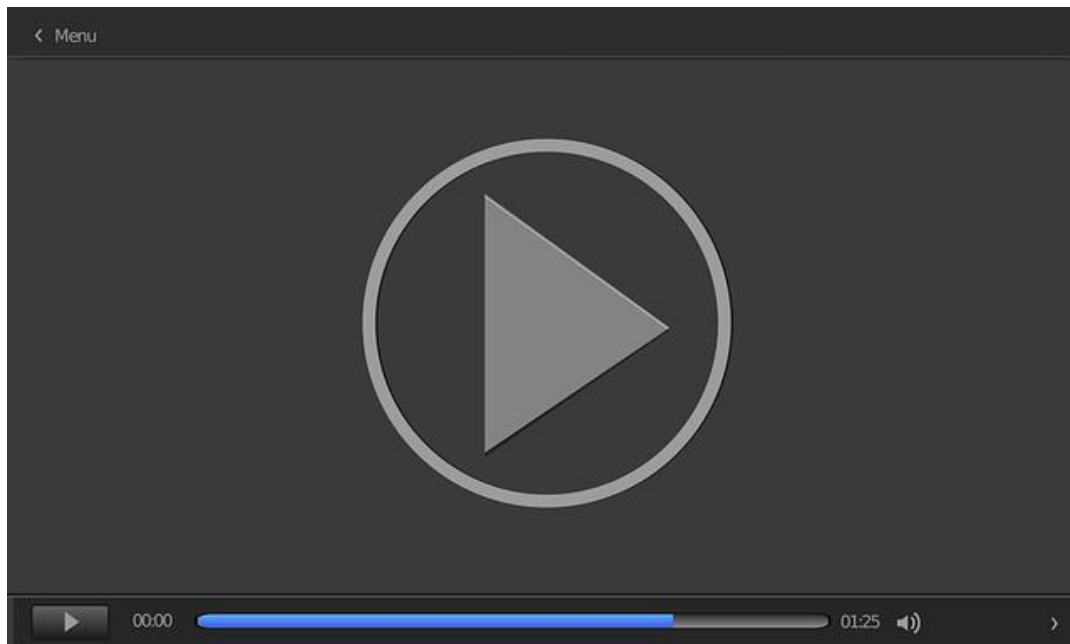


Figura 10. Fases para la construcción de un proyecto BI. Fuente: elaboración propia.

- ▶ Partimos de los objetivos estratégicos de la empresa.
- ▶ El estado actual hace referencia a si se está o no preparado para implementar una estrategia de BI. Puede que culturalmente estemos preparados; puede que haya empresas que estén muy evolucionadas en infraestructura, pero no en organización.
- ▶ En el estado deseado debemos hacernos la siguiente pregunta: ¿a dónde quiero llegar, según los objetivos estratégicos? Por ejemplo, yo quiero que todas las decisiones de la compañía se rijan por datos.
- ▶ En la estrategia BI debemos preguntarnos: ¿qué habilidades debo tener al interior de la organización?, ¿cómo promover el uso de las herramientas?, ¿cómo implementar el gobierno del dato? Si no definimos procesos de BI al interior de la compañía, no llegaremos a buen término.
- ▶ En el *roadmap* de soluciones, tenemos que buscar iniciativas para poder llegar al estado deseado. Priorizar iniciativas, de acuerdo con lo que la empresa tiene y puede.

- ▶ Diseño, construcción y evaluación son tareas asignadas al departamento de tecnología. Teniendo en cuenta los anteriores pasos, ya se podría pensar en qué diseño y qué herramientas son las más adecuadas.

A continuación, accede al vídeo *Roadmap del ingeniero de datos*.



Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=798e459f-8faa-4608-9b67-ad6600899165>

2.5. Business intelligence vs. business analytics

La gestión de la empresa está fundamentada en la toma de decisiones más apropiada para cumplir con los objetivos del negocio, satisfacer las necesidades de los clientes y empleados y mantener o mejorar la calidad de los productos. Con el avance en las tecnologías de información y las comunicaciones, el aumento en la capacidad de almacenamiento de datos ha dado paso a nuevas metodologías tales como: el *business intelligence* (BI) y el *business analytics* (BA), que ayudan y facilitan el proceso de toma de decisiones.

- ▶ **Business intelligence:** la inteligencia de negocios (BI) es un instrumento mediante el cual diferentes organizaciones pueden apoyar la toma de decisiones basadas en información precisa y oportuna para garantizar la generación del conocimiento necesario que permita seleccionar la alternativa que sea más conveniente para el éxito de la empresa (Rosado y Rico, 2010).
- ▶ **Business analytics:** es un conjunto de técnicas (entre las que se encuentran algoritmos predictivos y modelos estadísticos) que le permiten a la organización predecir posibles eventos o resultados. Esto es, se enfoca en el análisis futuro en función de la información de la empresa y modelos predictivos para apoyar la toma de decisiones y mejorar los procesos y, por ende, la competitividad del negocio (Thorlund y Laursen, 2017).

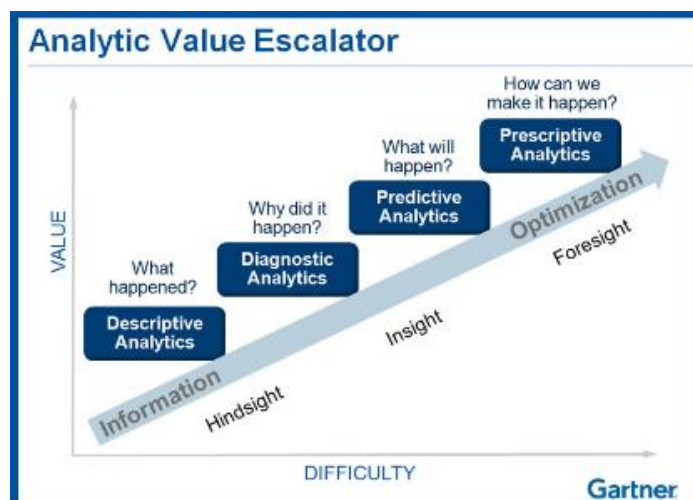


Figura 11. *Analytic value escalator*. Fuente: Pérez, s.f.

En resumen, se puede entender el *business intelligence* como las técnicas de **recoger** y **entender** datos del pasado, mientras que el *business analytics* permite alcanzar una visión más clara del futuro. Ambas metodologías se pueden complementar para construir un análisis minucioso de la actividad y futuro de la empresa, con el propósito de mejorar la toma de decisiones.

<i>Big data</i>	<i>Business intelligence</i>	<i>Business analytics</i>
Herramientas	Consultas, alertas, reportes, OLAP	Predicción, clasificación, regresión, agrupación
Centro	Qué, cómo paso y qué está pasando	Qué puede pasar
Uso	Reactivo	Proactivo, predictivo
Tipo de datos	Estructurado	Estructurados, no estructurados
Alcance	Dirección	Proceso

Figura 12. Diferencias entre *business intelligence*, *business analytics* y *big data*. Fuente: Curto, 2012.

Por otro lado, el *business analytics* es el análisis de las respuestas proporcionadas por el *business intelligence*. Mientras que el *business intelligence* responde a la pregunta «¿qué sucedió?», el *business analytics* responde a «¿por qué sucedió, volverá a pasar?». El *business intelligence* incluye informes, monitoreo automatizado y alertas, tableros y cuadros de mando integral; el *business analytics*, por el contrario, incluye análisis estadísticos cualitativos y cuantitativos, minería de datos, modelado predictivo y pruebas multivariantes. Cuando escuchas el término «inteligencia empresarial», normalmente engloba todo *business intelligence* y *business analytics*.

¿Qué beneficios tiene para la toma de decisiones usar el *business analytics*?

El conocimiento adquirido en *business analytics* permite a las organizaciones automatizar y optimizar sus procesos.

De hecho, las organizaciones impulsadas por datos que utilizan *business analytics* obtienen una ventaja competitiva porque pueden usar los conocimientos para (Thorlund y Laursen, 2017):

- ▶ Realizar minería de datos (explorar datos para encontrar nuevos patrones y relaciones).
- ▶ Realizar un análisis estadístico cualitativo y cuantitativo para explicar por qué ocurren ciertos resultados.
- ▶ Evaluar decisiones anteriores utilizando diferentes formas de análisis de datos.
- ▶ Utilizar el modelado predictivo y el análisis predictivo para pronosticar resultados futuros.

Business analytics también ofrece soporte a las organizaciones en el proceso de tomar decisiones tácticas proactivas y hace posible que esas organizaciones automaticen la toma de decisiones para respaldar las respuestas en tiempo real.

La construcción básica de *business analytics* se fundamenta en los cuatro pilares de Gartner (Popkin y Hayward, 2004), que se constituyen con cuatro componentes básicos:

- ▶ Los datos.
- ▶ Las personas.
- ▶ Los procesos.
- ▶ La tecnología.

Ayudan a recordar las cuatro áreas clave que se deben tener en cuenta al considerar la implementación del *business analytics*.

La intención detrás de estos pilares, como podemos observar en la Figura 13, es simplemente ayudar a hacer mejores preguntas y obtener mejores respuestas al desarrollar aplicaciones analíticas comerciales.

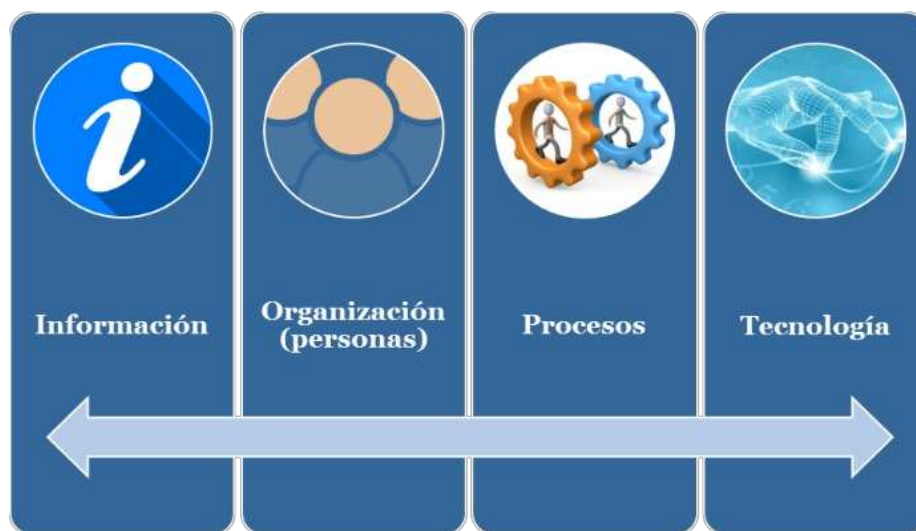


Figura 13. Pilares del *business analytics* según Gartner. Fuente: adaptado de Marrow, 2018.

- ▶ **Información:** el pilar de datos equilibra el manejo de la información. Requiere conectarse a fuentes de datos dispares (bases de datos, sistemas de información de

otras empresas y datos en la web, entre otros), independientemente de su tipo y ubicación.

Ser capaz de aprovechar todos los datos disponibles y servirlo al usuario es un componente básico crítico para una estrategia sólida. La información como base fundamental se ha agregado a la estructura para reflejar la conectividad y la coexistencia con todas las fuentes de datos que utiliza el análisis empresarial, no simplemente el almacén de datos. Esto se ha expandido para incorporar datos estructurados y no estructurados (contenido), datos locales y basados en la nube, y hemos visto surgir nuevos términos como *big data* para representar nuevos desafíos de información extrema, no solo de volumen, sino también de velocidad, variedad y complejidad de información.

- ▶ **Personas:** a medida que el *business analytics* se separa de un modelo centralizado hacia un modelo descentralizado, las personas deben estar capacitadas en saber cómo usar los datos. En este marco se ajustan las actividades de las personas para representar tareas en lugar de roles. Anteriormente, los roles eran los que mejor representaban la relación tradicional entre el negocio y las TI, las nuevas formas de análisis superan estas distinciones. Por lo tanto, un usuario de análisis empresarial puede participar fácilmente en producir, consumir y habilitar nuevas actividades que se conviertan en conocimiento.
- ▶ **Procesos:** el pilar del proceso requiere tener la información correcta en el momento adecuado para tomar mejores decisiones y más rápidas. Debido a que diferentes roles toman decisiones diferentes, es importante aprovechar los mismos datos para respaldar una variedad de procesos. Por ejemplo, los usuarios operativos y ejecutivos requieren cuadros de mando; los clientes o ciudadanos requieren declaraciones, propuestas e informes. Esto es para reforzar el punto de que el marco está compuesto por personas, procesos, plataformas y aspectos de desempeño.
- ▶ **Tecnología:** el pilar tecnológico abarca el desarrollo y la implementación de

sistemas que permitan desplegar todo el conocimiento. Las organizaciones deben construir una arquitectura flexible que se adapta a las necesidades del negocio.

Proceso de *business analytics*

El proceso completo de *business analytics* implica tres pasos principales aplicados secuencialmente a una fuente de datos (véase Figura 14). El resultado del proceso debe estar relacionado con las empresas, intentando mejorar el rendimiento constantemente.

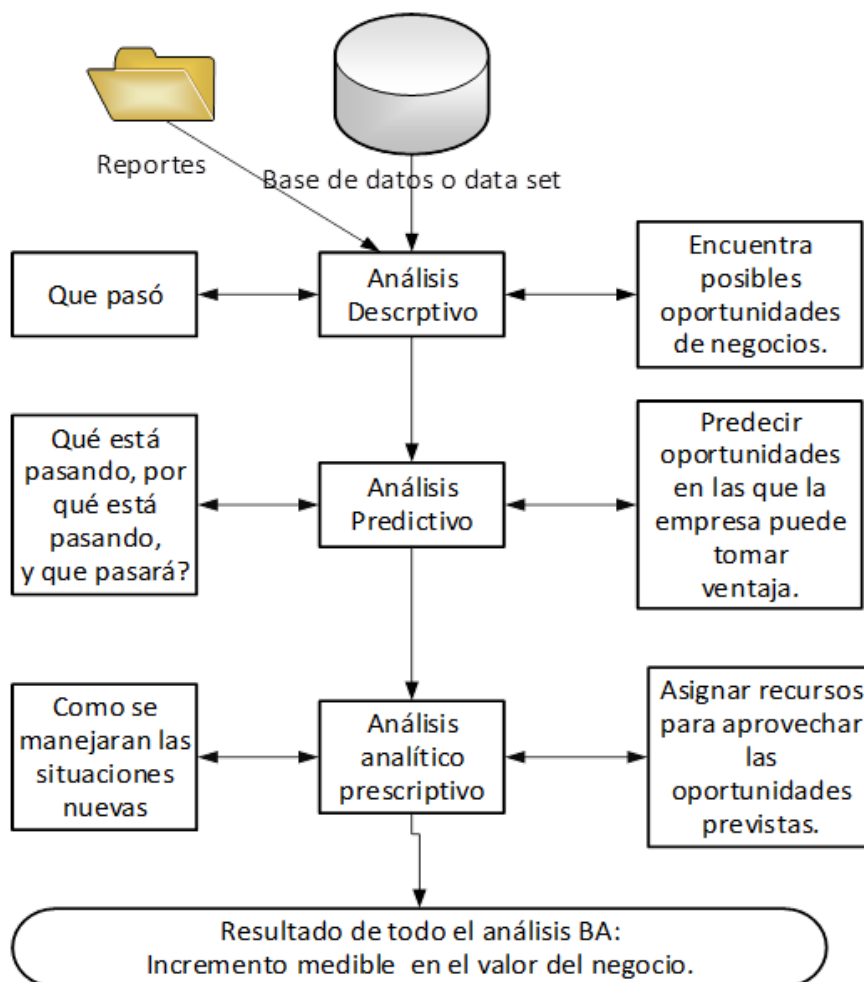


Figura 14. Proceso de *business analytics*. Fuente: Schniederjans et al., 2014.

La lógica del proceso en la Figura 14 se basa inicialmente en una pregunta: ¿qué

valiosa información está encerrada en las fuentes de datos que la organización tiene disponibles? En cada uno de los tres pasos que componen el proceso se deben responder las preguntas adicionales. Responder a todas requiere extraer la información de los datos a través de los tres pasos de análisis que comprenden el proceso.

El tamaño de algunas fuentes de datos puede ser inmanejable, demasiado complejo y generalmente confuso. La organización de los datos y el intento de dar sentido a su valor informativo requieren la aplicación de análisis descriptivos como primer paso en el proceso de *business analytics*. Uno puede comenzar simplemente clasificando los datos en grupos usando las cuatro clasificaciones posibles presentadas en la Figura 15.

Tipo de datos	Descripción
Datos categóricos	Datos que están agrupados por una o más características. Los datos categóricos usualmente involucran números cardinales contados o expresados como porcentajes. Ejemplo 1: mercados de productos que pueden caracterizarse por categorías de productos «con altos ingresos» o productos «de bajos ingresos», basados en ventas. Ejemplo 2: una encuesta donde se recoge información sobre variables como el género, estado civil o afiliación política. Es común usar este término para aplicar a conjuntos de datos que contienen elementos identificados por categorías, así como a observaciones resumidas en tabulaciones cruzadas o tablas de contingencia.
Datos ordinales	Datos clasificados u ordenados para mostrar preferencia relacional. Ejemplo 1: clasificaciones de equipos de fútbol no basadas en puntos anotados sino en victorias. Ejemplo 2: <i>ranking</i> de empresas comerciales basadas en la calidad del producto. Ejemplo 3: <i>ranking</i> de universidades teniendo en cuenta la calidad de sus investigaciones.
Datos de intervalo	Los datos que se organizan a lo largo de una escala donde cada valor es igualmente distante de los demás. Son datos ordinales. Ejemplo 1: un indicador de temperatura. Ejemplo 2: instrumento de encuesta que usa una escala Likert para medir la satisfacción del cliente (es decir, 1, 2, 3, 4, 5, 6, 7). De 1 a 2 se percibe como equidistante al intervalo de 2 a 3, y así sucesivamente. Nota: En los datos ordinales, la clasificación de las empresas puede variar mucho del primer lugar al segundo, pero en los datos de intervalo, deberían ser relacionamente proporcionales.
Ratios de datos	Datos expresados como una relación en una escala continua. Ejemplo 1: la proporción de empresas con programas de fabricación ecológica es el doble que la de empresas sin dicho programa.

Figura 15. Tipos de escalas de clasificación de medición de datos. Fuente: Schniederjans et al., 2014.

Además, también se pueden incorporar algunos de los datos en hojas de cálculo como las de Excel y preparar tabulaciones cruzadas y tablas de contingencia para restringir los datos a una estructura de datos más manejable. Se pueden calcular medidas simples de tendencia central y dispersión para intentar capturar posibles oportunidades de mejora de los procesos. Otros métodos descriptivos de resumen analítico, que incluyen trazado y gráficos, pueden ayudar a los responsables de la

toma de decisiones a visualizar los datos para comprender mejor las oportunidades de negocio.

Desde el **paso 1**, el **análisis descriptivo analítico** (véase Figura 14), algunos patrones o variables del comportamiento de la empresa se deben identificar para que representen los objetivos y las oportunidades de negocio, sumado al posible comportamiento futuro de las tendencias. Es probable que se requiera un esfuerzo adicional: la generación de informes estadísticos detallados estrechamente enfocados en los datos y relacionados con los objetivos del negocio para explicar lo que está ocurriendo (lo que sucedió en el pasado).

Esto es como una búsqueda estadística de variables predictivas en los datos que pueden conducir a encontrar patrones de comportamiento que una empresa podría aprovechar, si los patrones de comportamiento ocurren en el futuro. Por ejemplo, una empresa puede hallar en su información general de ventas que, durante los tiempos de inactividad económica, ciertos productos se venden a clientes de un nivel de ingresos específico y con una determinada publicidad. Las variables de ventas, clientes y publicidad pueden tener la forma de cualquiera de las escalas de datos descritas en la Figura 15 (Schniederjans et al., 2014).

Teniendo en cuenta los resultados del paso 1, se pueden determinar tendencias observadas y usarlas para pronosticar el futuro en el paso 2.

Paso 2: análisis predictivo del proceso de *business analytics*. Hay muchos métodos que se pueden emplear aquí. Una metodología comúnmente utilizada es la regresión múltiple. Esta es ideal para establecer si existe una relación estadística entre las variables predictivas encontradas en el análisis descriptivo. La relación podría mostrar que una variable dependiente se asocia de manera predictiva con el valor comercial o el rendimiento de algún producto.

Explorar las bases de datos de la empresa utilizando procedimientos estadísticos avanzados para verificar y confirmar las mejores variables predictivas es una parte

importante de este paso en el proceso. Esto responde a las preguntas sobre qué está sucediendo actualmente y por qué sucedió. Un modelo de regresión único o múltiple con frecuencia puede ayudar a pronosticar una línea de tendencia en el futuro. Cuando la regresión no aporta información, se pueden aplicar otros métodos de pronóstico (*exponential smoothing*, *smoothing averages*) como análisis predictivo para desarrollar los pronósticos necesarios de las tendencias del negocio (Schniederjans et al., 2014).

La **identificación de las tendencias futuras** es el resultado principal del paso 2. Esto ayuda a responder la pregunta «¿qué pasará?».

E n el **paso 3**, el **análisis prescriptivo**, las metodologías de investigación de operaciones se pueden utilizar para asignar de manera óptima los recursos limitados de una empresa y aprovechar al máximo las oportunidades que se encontraron en las tendencias futuras previstas. Los límites en recursos humanos, tecnológicos y financieros impiden que una empresa busque todas las oportunidades que puede tener disponibles en el momento. El uso del análisis prescriptivo le permite a la empresa asignar recursos limitados para alcanzar los objetivos de la mejor manera posible en el menor tiempo posible (Schniederjans et al., 2014).

En resumen, los tres componentes principales del proceso de BA (descriptivo, predictivo y prescriptivo) pueden ayudar a una empresa a encontrar oportunidades en sus datos, predecir tendencias que pronostican oportunidades futuras y ayudar a seleccionar una línea de acción que optimice la distribución de recursos de la empresa para maximizar el valor, el rendimiento y el desempeño.

2.6. Referencias bibliográficas

Anderson, C. (23 de junio de 2008). The end of theory: the data deluge makes the scientific method obsolete [Página web].

Wired. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Burton, B., Geishecker, L., Hostmann B, Friedman, T., y Newman, D. (2006). *Organizational Structure: Business Intelligence and Information Management*. Gartner.

Casey, T., Krishnamurthy, K., y Abezgauz, B. (12 de agosto de 2013). Who should own big data? [Página web]. *Strategy+Business*. <https://www.strategy-business.com/article/00211?gko=44b8e>

Curto, J. (2012). *Introducción al business intelligence*. UOC.

Dataworks. (s.f.). Integrity in the data lifecycle [Página web]. *Dataworks*. <https://www.dataworks.ie/5-stages-in-the-data-management-lifecycle-process/>

Davenport, T. H., Barth, P., y Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, 54, 22-24.

Debortoli, S., Müller, O., y Vom-Brocke, J. (2014). Comparing Business Intelligence and big data skills. *Business and Information Systems Engineering*, 6(5), 289-300.

Firincan, G. (8 de febrero de 2017). The 10 Vs of Big Data [Página web]. *TWDI*. <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx>

Gagliardi, E. O., Grosso, A., Turull, J. M., Piffaretti, P., y Pereyra, S. R. (1999). *Computación de queries a bases de datos relacionales utilizando circuitos booleanos* [Conferencia]. V Congreso Argentino de Ciencias de la Computación. San Luis,

Argentina.

Han, J., Kamber, M., y Pei, J. (2006). *Data mining, concepts and techniques*. Morgan Kaufmann.

Herrera, F., Riquelme, J., y Ruiz, R. (2004). Preprocesamiento de datos [Diapositivas]. Reunión Red Nacional DM & ML. <http://www.lsi.us.es/redmidas/IIreunion/trans/prepro.pdf>

IDS2015. (5 de abril de 2015). Inteligencia de Negocios (BI) [Mensaje en un blog]. *Ingeniería del software U A H*. <https://ingenieriadelsoftwareuah2015.wordpress.com/2015/04/05/inteligencia-de-negocios/>

Kimble, C. (2013). Knowledge management, codification and tacit knowledge. *IR Information Research*, 18(2). <http://informationr.net/ir/18-2/paper577.html>

Lin, T. Y. (2002). Attribute transformation for data mining I: theoretical explorations. *International Journal of Intelligent Systems*, 17, 213-222.

Marrow, G. (2018). *Business analytics & intelligence. An introduction and considerations for getting started*. Durham County Government.

McAfee, A., y Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 61-67.

Pérez, M. F. (s.f.). Transformación digital en la empresa: análisis predictivo [Página web]. *Sistel*. <https://www.sistel.es/transformacion-digital-empresa-analisis-predictivo>

Popkin, J., y Hayward, B. (2004). *Top 10 strategic technologies for 2005*. Gartner Symposium/ITxpo.

Puklaveč, B., Oliveira, T., y Popovič, A. (2018). Understanding the determinants of business intelligence system adoption stages. *Industrial Management and Data*

Systems, 118(1), 236-261.

Pyle, D. (1999). *Data Preparation for data mining (the Morgan Kaufmann series in data management systems)*. Morgan Kaufmann.

Rosado, A. A., y Rico, D. W. (2010). Inteligencia de negocios: estado del arte. *Scientia et Technica*, 16(44), 321-326. <http://revistas.utp.edu.co/index.php/revistaciencia/article/view/1803/1209>

Schniederjans, M. J., Schniederjans, D. G., y Starkey, C. M. (2014). *Business analytics principles, concepts and applications: What, why, and how*. Pearson Education.

Sharma, S., Sharma, J., y Devi, A. (2009). Corporate social responsibility: the key role of human resource management. *Business Intelligence Journal*, 2(1), 205-213. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.514.7758&rep=rep1&type=pdf>

Teixeira, A., Oliveira, T., y Varajão, J. (2019). Evaluation of business intelligence projects success - a case study. *Business Systems Research*, 10(1) 1-12.

Thorlund, J., y Laursen, G. H. N. (2017). *Business analytics for managers*. Wiley.

Trejo, D. (2020). *Gobierno de datos para directores: Realizando la transformación digital a partir de los datos*.

Valenzuela, L. M. (2007). *La gestión del valor de la cartera de clientes y su efecto en el valor global de la empresa: diseño de un modelo explicativo como una herramienta para la toma de decisiones estratégicas de marketing* (Tesis doctoral). Universidad Complutense de Madrid. <http://eprints.ucm.es/8064/1/T29976.pdf>

Weller, K. (2010). *Knowledge representation in the social semantic web*. De Gruyter.

Introducción al business intelligence

Curto, J. (2012). *Introducción al business intelligence*. UOC.
<https://docplayer.es/709686-Introduccion-al-business-intelligence.html>

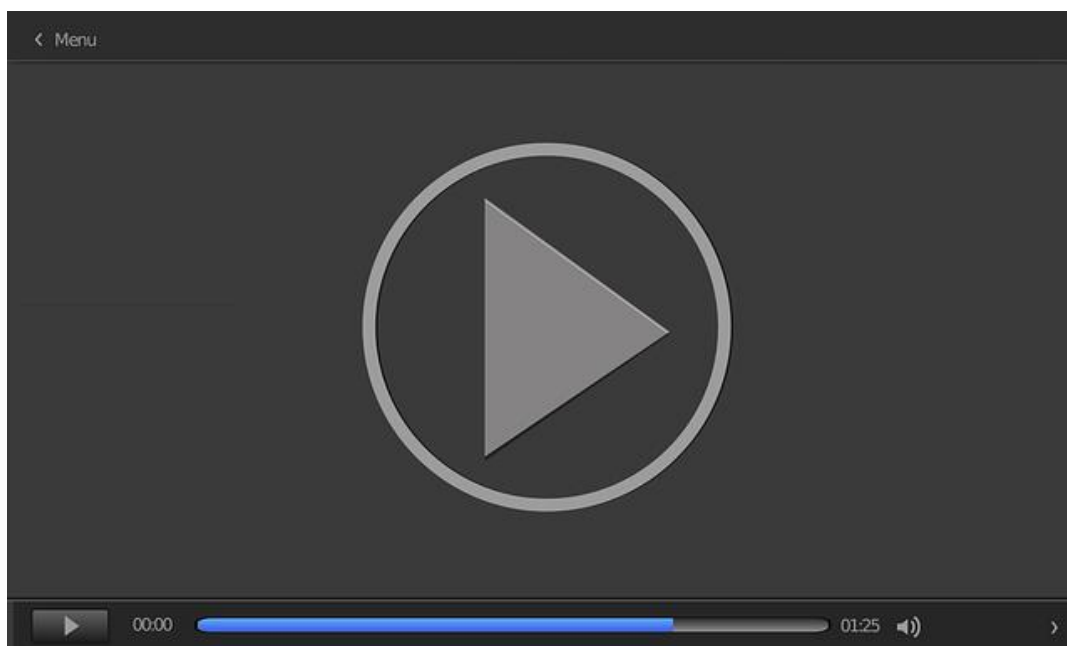
El *business intelligence* (o inteligencia de negocio) es un concepto complejo. No por su definición, que es sencilla de enunciar y comprender, sino principalmente por el hecho de que en él confluyen una gran cantidad de tecnologías, metodologías, procesos y estrategias que complican sobremanera la iniciación al neófito. Además, a lo largo del ciclo de vida de estos sistemas de información, se incrementa la complejidad de la arquitectura, así como las necesidades de negocio y las tecnologías que las soportan.

Sin embargo, la gran mayoría de organizaciones necesita actualmente este tipo de sistemas de información para tomar mejores decisiones y ser más competitivas. El *business intelligence* se convierte en una de las principales necesidades. Por ello es necesario poder construir soluciones sólidas a partir de conocimientos profundamente asentados. Este libro introduce los principales conceptos de la inteligencia de negocio a través de las principales fases de diseño de un proyecto de este tipo para constituir una sólida base de adquisición de conocimientos más profundos.

Business intelligence y business analytics

AddKw. (3 de abril de 2015). *Business Intelligence VS Business Analytics* [Video]. Youtube. https://www.youtube.com/watch?v=gGtb_FpJHTk

La organización sin ánimo de lucro LPI ADDKW explica de qué forma aplica tanto la inteligencia como la analítica de negocios y por qué estas pueden resultar útiles en cualquier otro organismo, ya sea gubernamental, empresarial, etc.



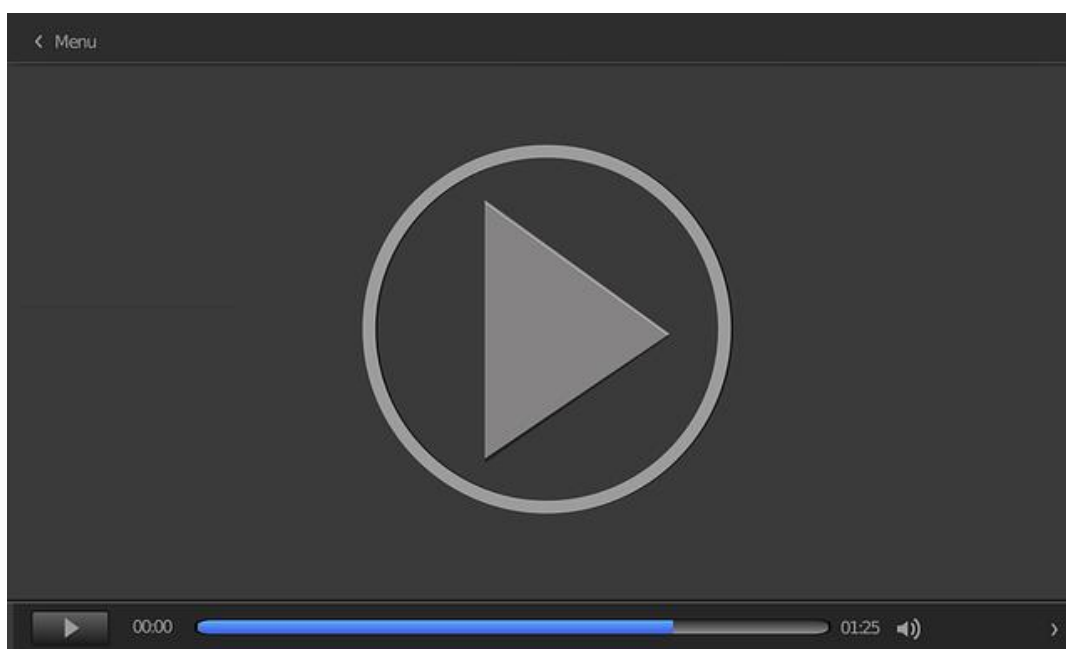
Accede al vídeo:

https://www.youtube.com/embed/gGtb_FpJHTk?si=Yzm9pVKGN6EVm2aE

Curso introductorio de business intelligence y business analytics

Consultora Re-Ingenia S. A. (23 de mayo de 2018). *Curso Introductorio Business Intelligence y Business Analytics* [Vídeo]. Youtube. https://www.youtube.com/watch?v=_AOic35nLvM

La consultora Re-Ingenia S. A. tiene a disposición un curso completo para aplicar los conceptos vistos en este temario. Tal y como afirma: «*business intelligence* es la metodología técnica para transformar los datos en información y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios».



Accede al vídeo:

https://www.youtube.com/embed/_AOic35nLvM?si=x_XYCKzByQliT1YE

PowerData

PowerData. Página web oficial. <https://www.powerdata.es/data-governance>

Presenta diferentes artículos en español donde presentan temáticas tales como *data governance*, calidad de los datos, MDM: ¿qué es y cómo debes implementarlo en tu empresa? y *data lake*, entre otros.

1. Elige la respuesta correcta.
 - A. Los datos no tienen relevancia, pues no tienen ningún significado si no se contextualizan.
 - B. Si el dato está bien contextualizado, obtenemos información veraz y confiable.
 - C. El dato es una pieza fundamental en las empresas hoy en día, si los datos cumplen un ciclo de vida y se tienen maestros de datos, se garantiza el éxito en los procesos BI y BA.
 - D. La información sin ser procesada genera conocimiento.

2. El ciclo de vida del dato en el gobierno de datos incluye:
 - A. Creación, uso y destrucción.
 - B. Almacenaje, distribución y archivo.
 - C. Creación, distribución y destrucción.
 - D. A y B son correctas.

3. No forma parte de los pasos importantes para construir un ecosistema BI.
 - A. Entender y tener en cuenta los objetivos estratégicos es fundamental para poder dar respuesta a lo que la organización se traza como meta.
 - B. Evaluar el estado actual y el estado deseado en cuanto a madurez de la organización.
 - C. Definir una estrategia BI, definir una ruta y diseñar y construir las aplicaciones.
 - D. Crear modelos de predicción.

4. ¿Cuáles son los cuatro pilares de Gartner para la construcción básica de business analytics?
 - A. Proceso, clientes.
 - B. Información, personas.
 - C. Procesos, tecnologías.
 - D. B y C son correctas.

5. Menciona dos pasos del proceso de business analytics:
 - A. Predictivo, analítico.
 - B. Descriptivo, transformador.
 - C. Descriptivo, predictivo.
 - D. Ninguna es correcta.

6. ¿Por qué es importante crear datos maestros?
 - A. Evita duplicidad de datos.
 - B. Aumenta la confiabilidad de los datos.
 - C. A y B son correctas.
 - D. A y B son incorrectas.

7. Menciona una diferencia del business analytics frente al business intelligence.
 - A. En el BA se realiza análisis predictivo y prescriptivo, en el BI se analizan datos pasados.
 - B. Uso de bases de datos NoSQL.
 - C. Uso de administración de bases de datos.
 - D. Integración y divulgación.

8. Identifica la afirmación falsa con respecto a los beneficios del business analytics:
- A. Las organizaciones obtienen ventaja competitiva.
 - B. Se puede realizar minería de datos para detectar patrones entre datos.
 - C. Se pueden simular diferentes escenarios y crear estrategias para mitigar riesgos.
 - D. Se puede realizar un análisis descriptivo y diagnóstico de los datos.
9. El análisis que da respuesta a la pregunta de cómo se manejarán futuros escenarios es:
- A. Análisis descriptivo.
 - B. Análisis predictivo.
 - C. Análisis prescriptivo.
 - D. Ninguna es correcta.
10. El business intelligence es:
- A. Reactivo.
 - B. Proactivo.
 - C. Predictivo.
 - D. Ninguno de los anteriores.