

Explotación de datos

Carles Llorach Rius

PID_00236086



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

Introducción	5
Objetivos	6
1. Beneficios de una adecuada presentación de datos	7
2. Consideraciones para la presentación de datos (riesgos)	11
3. Formatos de presentación	17
3.1. Informes	17
3.2. Análisis OLAP	18
3.3. Cuadros de mando	19
3.4. Otros	21
3.4.1. Sistemas de apoyo a la toma de decisiones (<i>decision support systems, DSS</i>)	21
3.4.2. Mapas	22
3.4.3. Minería de datos (<i>data mining</i>)	23
3.4.4. Autoservicio BI	24
3.4.5. Sistemas de búsqueda empresarial en lenguaje natural	25
3.4.6. <i>Big data</i>	26
3.4.7. <i>Webhousing y mobile BI</i>	26
4. Herramientas de apoyo a la presentación de datos	28
4.1. Metodologías	28
4.2. Técnicas y componentes	31
4.3. Recursos en línea	33
4.4. Herramientas de visualización	36
4.5. Herramientas de apoyo	39
Glosario	41
Bibliografía	43

Introducción

Los sistemas de almacenes de datos son una de las iniciativas más robustas que se pueden emplear para ayudar a las organizaciones y, más específicamente, a sus accionistas a crear más valor. Su éxito se basa en la explotación de los datos que provee el sistema, de tal modo que se conviertan en una ventaja sobre sus competidores y permitan resolver problemas al unir a personas y tecnología.

La visualización de datos es una preocupación clave para las organizaciones y los profesionales de la inteligencia de negocio y analistas, ya que afecta de forma determinante al valor que se obtiene de la información.

Cuando las condiciones de visualización de datos son las adecuadas, los usuarios pueden interactuar con los datos de manera mucho más efectiva. El resultado es el enriquecimiento de la organización, que disfruta de un nivel de conocimiento como nunca antes y se encuentra mejor posicionada a la hora de tomar decisiones e incrementar la productividad.

Más allá de los instrumentos tradicionales que viene ofreciendo el BI (*reporting*, análisis multidimensional y cuadros de mando), encontramos nuevos instrumentos (mapas, análisis exploratorio, minería de datos, etc.) En todos ellos, el descubrimiento de datos visual e interactivo ha contribuido a acelerar la transición hacia el autoservicio de BI gracias a su fácil configuración, que lo hace aún más fácil de usar.

Los retos a los que se enfrenta la visualización de datos son importantes, mucho más en la era actual del *big data*, cuando un error puede magnificar sus consecuencias de manera irremediable.

Afortunadamente, disponemos de una amplia paleta de recursos: metodologías, herramientas, guías, etc. que nos van a facilitar la publicación de nuestros datos, de tal manera que su visualización sea la más adecuada para el destinatario de esta información.

Objetivos

Los contenidos incluidos en este módulo se orientan a conseguir que el estudiante alcance los objetivos siguientes:

- 1.** Tomar conciencia de la importancia de la visualización de datos en condiciones óptimas para mejorar la toma de decisiones, fomentar el intercambio de información y mejorar la comunicación.
- 2.** Comprender la utilidad e idoneidad de los distintos formatos de presentación existentes.
- 3.** Conocer el amplio conjunto de recursos en línea disponibles y los principios que deben regir su utilización.
- 4.** Conocer las herramientas de soporte y presentación de datos existentes y estar capacitados para evaluarlas frente a las necesidades de la organización.
- 5.** Identificar los riesgos que supone la explotación de datos de baja calidad y también su inadecuada visualización.

1. Beneficios de una adecuada presentación de datos

Un adecuado análisis de los datos existentes en una empresa u organización puede facilitar enormemente la comprensión de los negocios y mercados, y es de gran ayuda para la toma de las decisiones empresariales correctas. Para llevar a cabo este análisis, es necesaria la utilización de las herramientas precisas que incrementen la eficiencia organizacional y su efectividad, agilizando el flujo de datos dentro de la organización.

La visualización de datos en condiciones óptimas fomenta el intercambio de información y mejora la comunicación. Además, otra particularidad de las visualizaciones es que transmiten la información de una manera universal, consiguiendo que resulte muy sencillo explicar algo y que además se pueda compartir con otros miembros del equipo o con integrantes de la organización en diferentes ámbitos.

Sin duda alguna, la cultura organizativa y el nivel de madurez en el que se encuentre una organización determinarán qué instrumentos se implantarán y cuándo:

- Si no hay una cultura de la medición, el proyecto debe fundamentarse en el *reporting* operacional, en el *reporting* táctico y en análisis multidimensionales controlados.
- Si la cultura de la medición está presente pero no arraigada, debe darse paso a los cuadros de mando tácticos, fomentando que los mandos intermedios comiencen a utilizar métricas de control.
- Si hay una cultura de la medición muy arraigada, el proyecto debe centrarse en madurar el *reporting*, el análisis y los cuadros de mando estratégicos, y empezar con el descubrimiento y la minería de datos.

También debemos plantearnos dentro del proyecto la definición y consolidación de la «verdad corporativa»: es necesario definir unos informes estándar que sean los oficiales y unos caminos de análisis básicos sobre estos informes, de manera que todo el mundo sepa dónde y cómo se debe acceder a la información.

Otro punto que hay que tener en cuenta es la uniformidad de la organización semántica de la información.

Por ejemplo, si se pregunta a varios usuarios qué entienden por margen de la organización, y obtenemos distintas respuestas (uno lo define antes de aplicar impuestos, otro después, y otro indica que hay que descontar los abonos), entonces deberemos crear un documento de semántica común antes de empezar el proyecto. Este documento deberá explicar «qué se entiende por...» y, además, hay que asegurar que todos los implicados utilicen la misma semántica.

Con los datos de un sistema BI, es posible generar reportes globales o por secciones; crear escenarios con respecto a una decisión; hacer pronósticos; análisis multidimensionales; generar y procesar datos, etc.

A continuación, se enumeran algunos ejemplos de las áreas más comunes en las que se utilizan las soluciones de inteligencia de negocio:

- Ventas: análisis de ventas; detección de clientes importantes; análisis de productos, líneas, mercados; pronósticos y proyecciones.
- Marketing: segmentación y análisis de clientes; seguimiento a nuevos productos.
- Finanzas: análisis de gastos e ingresos; rotación de cartera; razones financieras.
- Manufactura: productividad en líneas; análisis de desperdicios; análisis de calidad; rotación de inventarios y partes críticas.
- Embarques: seguimiento de embarques; motivos por los cuales se pierden pedidos.

Encontramos una gran diversidad de enfoques y prácticas que varían de un país a otro, pero es universalmente aceptado el hecho de que una de las mejores técnicas para hacer comprensibles los datos es la representación de los números mediante imágenes. Esto puede hacer mucho más fácil apreciar un patrón o exponer ciertos patrones que, de otro modo, podrían quedar ocultos.

La información a la que acceden los usuarios resulta del tratamiento y presentación adecuados de los datos para la toma de decisiones a tres niveles:

1) **Nivel operativo:** permite a los usuarios de este nivel, que en su rutina diaria manejan información, que la reciban de una manera oportuna, exacta y adecuada.

Manejan principalmente herramientas de reportes u hojas de cálculo con formatos fijos que están en constante actualización.

Pongamos el ejemplo de un supervisor de ventas cuyo principal apoyo para hacer satisfactoriamente su trabajo es una hoja de cálculo, con la cual monitorizaría el hecho de que se cumplan las cuotas de ventas de los vendedores a su cargo. Una de las columnas tendría información fija (cuota de ventas), y en la siguiente se anotarían los datos diarios de ventas (en términos monetarios) de cada uno de sus vendedores, y con estos datos podría hacer un análisis y tomar decisiones respecto a los objetivos que cumple cada trabajador o, en su caso, sus deficiencias.

2) Nivel táctico: este nivel permite analizar y consultar información en un nivel medio, mediante la utilización de herramientas informáticas, sin intervención de terceros.

Supongamos, por ejemplo, que a un gerente le llega un informe impreso, o preimpreso, donde indica que las ventas en algún producto o servicio se han elevado considerablemente.

Una herramienta de este nivel debe permitir llevar a cabo un análisis que marque las tendencias y variables de un posible incremento. Gracias a la misma, el gerente puede darse cuenta de que el incremento de demanda se debe a productos, clientes o estrategias promocionales. Además, dicha herramienta puede determinar si es cíclico (y ocurre en un periodo definido y su tendencia aumenta en una época determinada) o si estos incrementos son puntuales.

A partir del análisis, el gerente puede dictar estrategias para aumentar el impacto positivo o reducir el impacto negativo, según el caso.

3) Nivel estratégico: permite monitorizar y analizar las tendencias, patrones, metas y objetivos estratégicos de la alta dirección.

En este nivel, es muy común encontrar cuadros de mando (*balanced scorecard*, término introducido por Robert Kaplan y David Norton). Tiene concordancia con normas internacionales de excelencia en la gestión y aseguramiento de la calidad.

Es un esquema con múltiples dimensiones para describir, implementar y administrar una estrategia en cualquier nivel, mediante la relación de objetivos, iniciativas y mediciones a las estrategias organizacionales. Conducen a la empresa a estrategias definidas que se convierten en planes de acción.

En la figura 1 se muestra para cada nivel de la toma de decisiones, las herramientas que se utilizan y los usuarios que requieren la información.



Figura 1. Herramientas, usuarios y niveles de toma de decisiones

(Fuente: www.gestiopolis.com)

Existen múltiples herramientas para cada nivel, y se seleccionan a partir de las necesidades de cada organización. Vemos en la figura 2 los perfiles de usuario de cada una de ellas, así como su nivel de funcionalidad:

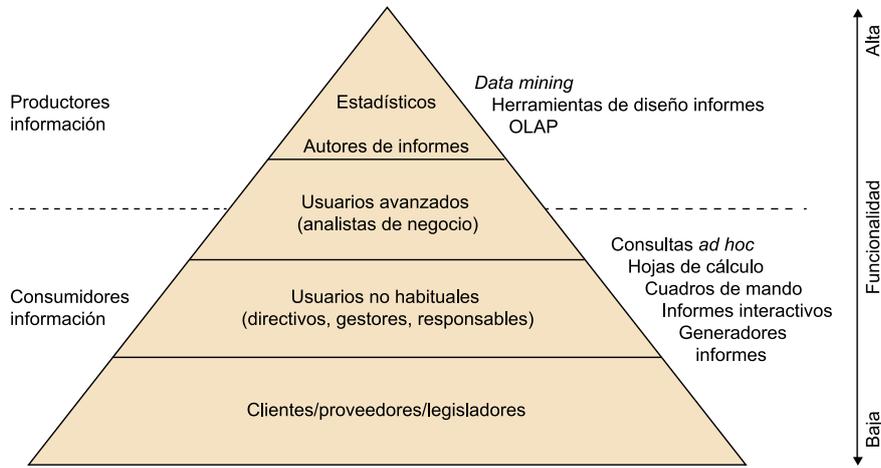


Figura 2. Herramientas y perfiles de usuarios
 (Fuente: J. L. Cano. *BUSINESS INTELLIGENCE: Competir con información*)

2. Consideraciones para la presentación de datos (riesgos)

Las organizaciones son conscientes de los problemas que les ocasiona la baja calidad de los datos, pero no siempre se contemplan los inconvenientes de una presentación de los datos inapropiada.

En la encuesta *Data Quality and the Bottom Line* (W. Eckerson, 2002), cuando se les preguntaba cuáles eran estos problemas, afirmaron:

- Tiempo extra para reconciliar los datos (87 %).
- Pérdida de credibilidad en el sistema (81 %).
- Insatisfacción de clientes (67 %).
- Retrasos en el desarrollo de nuevos sistemas (64 %).
- Pérdidas de ingresos (54 %).
- Problemas de conformidad (38 %).
- Etc.

Y cuando se les preguntaba por los beneficios que aporta la adecuada visualización de datos, afirmaban:

- Simple versión de la verdad (19 %).
- Incrementos en la satisfacción de los clientes (19 %).
- Mayor confianza en los sistemas de análisis (17 %).
- Reducción de costes (13 %).
- Menor tiempo para reconciliar los datos (12 %).
- Incremento de ingresos (9 %).
- Etc.

En las organizaciones, nos encontramos con un ecosistema de tecnologías que son vitales para su normal funcionamiento y que maneja datos complejos y dinámicos. En muchas ocasiones, estas tecnologías son gestionadas de manera individual, sin abordar una gestión corporativa conjunta, que resultaría sin duda más eficiente.

Veamos las principales ventajas que aportan las herramientas de visualización y análisis:

- Automatización de grandes volúmenes de datos: trabajar manualmente los datos sería inexacto e incluso arriesgado.
- Integración: destacan, sobre todo, sus funciones de integración con otras aplicaciones específicas que utilice la empresa.
- Accesibilidad: permite a los usuarios trabajar en cualquier momento, desde cualquier lugar y dispositivo.
- Usabilidad: interfaz intuitiva, interactiva, integrada y unificada.
- Alto rendimiento en el procesamiento de datos, tanto en paralelo como con escalabilidad.
- Disminuye el tiempo de elaboración, de modo que la alta gerencia gane tiempo para el análisis.
- Información relevante actualizada.

Si se dispone de las herramientas adecuadas con el personal inadecuado y/ o poco capacitado, la empresa corre el riesgo de fracasar en el propósito de publicar adecuadamente los datos, debido al mal uso de la tecnología.

Existen también otras circunstancias técnicas que no contribuyen a la correcta visualización de datos, y que conviene corregir o evitar siempre que sea posible:

- La tecnología evoluciona rápidamente, La rápida evolución de las tecnologías lleva a introducir nuevas herramientas en la empresa, con lo que se corre el riesgo de que estas no sean utilizadas o se empleen de manera inadecuada hasta ser comprendidas y dominadas por los usuarios.
- Acceso erróneo de los datos.
- El hecho de que la empresa tenga múltiples herramientas de BI, con datos poco claros y metas incompatibles con la organización, induce a tener diferentes conclusiones de los mismos datos.

- No llega a abarcar todo el conjunto de procesos de la inteligencia de negocios de la organización.
- Instalación compleja y problemas de rendimiento.
- Ante un conjunto de problemas, ver soluciones aisladas por departamento y no como partes de un mismo sistema.
- Se desconoce el impacto que la mala interpretación de los datos, escasa información o la sobreinformación pueden ocasionar.
- No existen antecedentes de análisis, o sus procesos no están debidamente estandarizados.

Las representaciones deben ilustrar las tendencias y las relaciones de manera rápida y sencilla. Son una forma eficiente de transmitir la información desde la base de datos a la cabeza del lector.

Sin embargo, hay que tener cuidado: una mala representación de la información puede resultar engañosa. Hay muchas maneras de proporcionar información engañosa, bien sea deliberadamente o, como ocurre de manera más frecuente, inintencionadamente.

Debe haber un equilibrio entre diseño y función, pues las representaciones complicadas a menudo no logran hacerse entender. Dado que la interpretación de gráficos puede ser complicada, no hay que forzar a los lectores a tener que desenmarañar el mensaje.

Nuestra capacidad de hacer observaciones visuales de manera rápida y fácil se basa en la capacidad del cerebro para percibir regularidades e irregularidades. Gran parte de esta capacidad funciona de manera inconsciente, ya que la comparación se produce casi antes de empezar a pensar en ello.

Por ejemplo: ¿cuál de los círculos es mayor, el del centro del diagrama de la izquierda o el del centro del diagrama de la derecha? Los círculos en el centro de cada diagrama son del mismo tamaño.

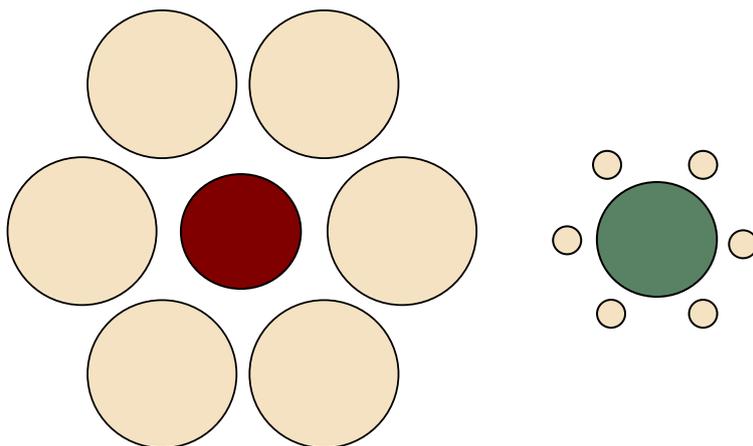


Figura 3. *The Ebbinghaus Illusion*
(Fuente: H. Ebbinghaus, 1850-1909)

Los malentendidos y malas interpretaciones también pueden ser resultado de diferentes tradiciones culturales. Los colores, por ejemplo, pueden tener diferente interpretación en distintas partes del mundo.

La **experiencia** también juega un papel en cómo se perciben los gráficos. Conviene conocer a los destinatarios de los datos, sus habilidades, experiencias y sus posibles diferencias. No debe asumirse que todo el mundo sabe lo mismo sobre un tema en cuestión.

A continuación, se exponen algunos aspectos que hay que considerar para maximizar la eficiencia de un gráfico:

1) Los datos deben ocupar un **lugar** central. Los componentes de apoyo deben:

- Ser expuestos solo si procede. Títulos de ejes, leyendas y etiquetas de datos pueden ser esenciales para la correcta comprensión del gráfico o no ser en absoluto necesarios, dependiendo de la naturaleza de los datos.
- Ser sutiles. Utiliza líneas más claras para los ejes y líneas de división que para los componentes de datos. Los elementos decorativos no tienen que distraer la atención del lector.

Veamos a continuación, en la figura 4, un ejemplo:

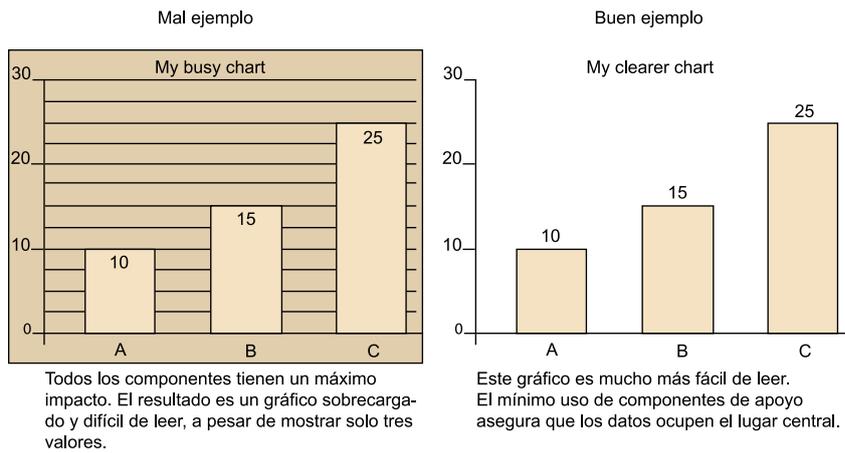


Figura 4. Ejemplos de distintas ubicaciones del dato (Fuente: Naciones Unidas. *Cómo hacer comprensibles los datos.*)

2) Al diseñar un gráfico, es posible ajustar las escalas para la mejor transmisión del mensaje.

Por ejemplo: los dos gráficos de líneas que se muestran en la figura 5 presentan los mismos datos, pero proporcionan imágenes muy diferentes:

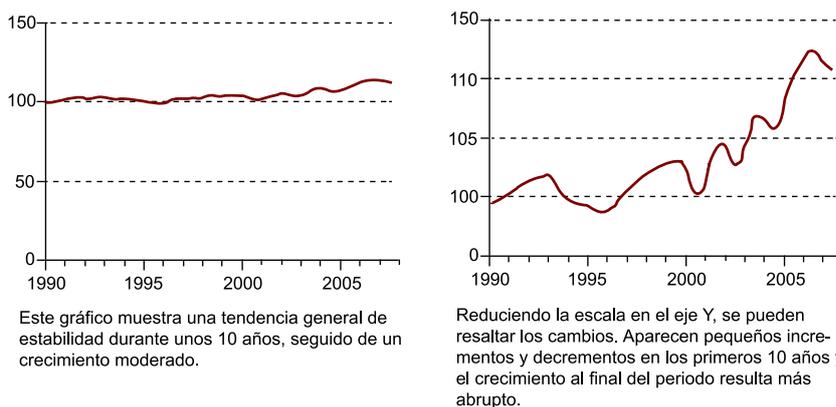


Figura 5. Ejemplos de distintas escalas de datos (Fuente: Naciones Unidas. *Cómo hacer comprensibles los datos.*)

3) Los datos pueden contener varios mensajes para resaltar en un gráfico. A los gráficos, al igual que a todos los demás elementos de una publicación, se les puede asignar una carga cognitiva.

La **carga cognitiva** significa básicamente cuánto tiene que esforzarse el lector para entender lo que se está tratando de comunicar.

Un gráfico con una elevada carga cognitiva será difícil de entender y recordar, y su mensaje resultará difícil de comunicar. Un gráfico con una reducida carga cognitiva se entiende con un vistazo, y su mensaje resulta obvio. La mayoría de las directrices sobre el diseño de gráficos eficaces están dedicadas a mantener una carga cognitiva baja.

A continuación, en la figura 6, se muestra un buen ejemplo de gráfico con carga cognitiva alta:

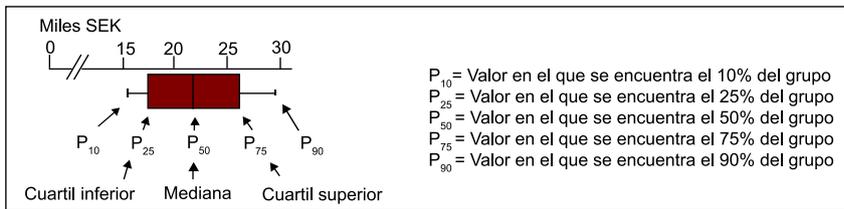
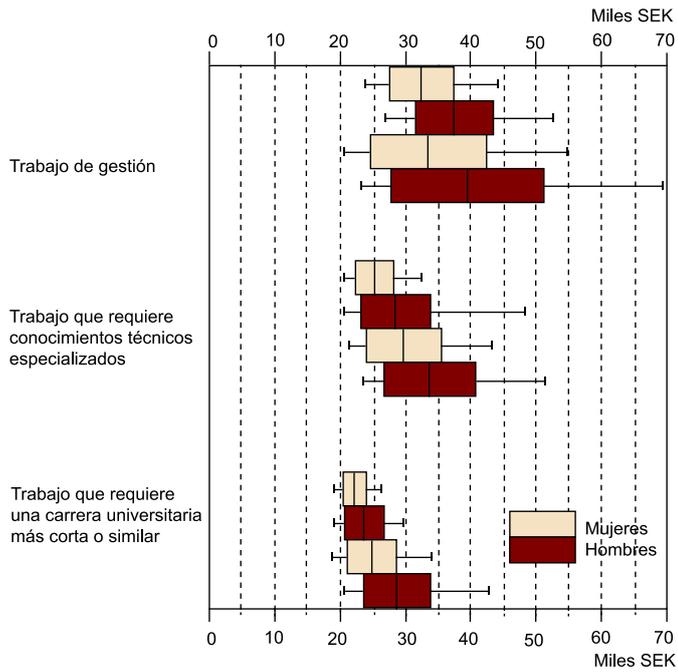


Figura 6. Ejemplos de carga cognitiva de un gráfico
 (Fuente: Naciones Unidas. *Cómo hacer comprensibles los datos.*)

Para resumir, las visualizaciones mal creadas pueden tener efectos muy negativos para el negocio, llegando a:

- Confundir a los usuarios.
- Dificultar la comprensión de los datos.
- Complicar el procesamiento de los datos diarios.
- Hacer perder la confianza en los sistemas de inteligencia de negocios.

3. Formatos de presentación

Es el área donde más avances se han producido en los últimos años. Sin embargo, la proliferación de soluciones mágicas y su aplicación coyuntural para solucionar aspectos puntuales ha llevado, en ocasiones, a una situación de desánimo en la organización respecto a los beneficios de una solución BI. Sin entrar a detallar las múltiples soluciones que ofrece el mercado, a continuación se identifican los modelos de funcionalidad o herramientas básicas (cada producto de mercado integra, combina, potencia, adapta y personaliza estas funciones).

Según el proceso que abarcan y el ámbito al que se orientan, las herramientas de *business intelligence* pueden clasificarse en:

- **Informes** (*reporting*): la utilidad de estas herramientas es generar documentos e informes con un alto nivel de detalle. Permiten la obtención de información actualizada de manera muy rápida y ágil, y la generación automática de alarmas a partir de unos criterios programados con anterioridad.
- **Análisis OLAP** (*on-line analytical processing*): generación de análisis muy completos con un enfoque deductivo y con la posibilidad de aplicar filtros personalizados.
- **Cuadros de mandos** (*dashboard* o *scorecard*): elaboración automática de diagramas y gráficos de gran poder visual, ya que permiten entender procesos y comparar datos con un simple vistazo.
- **Otras**: hojas de cálculo, herramientas de exploración, sistemas geográficos, minería de datos, etc.

3.1. Informes

El *reporting* es el precursor de la toma de decisiones. Al examinar un informe, se debe tener la mente abierta para explorar el contenido y profundizar en lo que se expresa con cifras y porcentajes. Las visualizaciones ayudan a los usuarios a darse cuenta de realidades que no eran obvias para ellos antes.

La principal ventaja de emplear un modo visual para exponer una determinada información es que, incluso cuando los volúmenes de datos son muy grandes, los patrones se pueden observar de manera rápida y sencilla.

Las herramientas de *reporting* existen desde hace mucho tiempo, y son soluciones maduras que permiten cubrir todas las necesidades de los usuarios finales, si bien en función de cada fabricante hay sutiles diferencias. Las tendencias actuales para estas herramientas incluyen la incorporación de mayores capacidades de visualización, autoaprovisionamiento y funcionalidad para embeber los informes en documentos (PDF, Word, PPT) u otras aplicaciones de negocio. También ofrecen cierta interacción, en informes que contienen tablas de datos que pueden ser enlazadas con otros informes. Esto implica que los usuarios puedan observar de manera resumida los datos dentro de una tabla de fácil comprensión y entonces, si les interesa, acceder a la información detallada de los mismos.

El uso de herramientas informáticas supone un impacto considerable en el rendimiento, no solo de la generación de los informes sino sobre todo de la operativa de negocio.

Se entiende por **plataforma de *reporting*** aquellas soluciones que permiten diseñar y gestionar (distribuir, planificar y administrar) informes en el contexto de una organización o en una de sus áreas.

Un informe permite responder principalmente a la pregunta de: ¿qué pasó? Y está destinado a los usuarios de negocio que tienen la necesidad de conocer la información consolidada y agregada para la toma de decisiones.

Por último, no podemos olvidar que los informes requieren un mantenimiento, en caso de que cambien los requisitos y los datos, y normalmente necesitaremos un control detallado del aspecto.

3.2. Análisis OLAP

Los procesos OLAP se han estudiado en profundidad en el módulo «Diseño e implementación multidimensional de un *data warehouse*» de la asignatura «Diseño y construcción del almacén de datos», pero debido a su gran importancia, merece la pena dedicar un pequeño apartado para recordar qué son.

Las herramientas OLAP permiten acceder a los datos de negocio de forma rápida, consistente e interactiva, teniendo en cuenta las diferentes perspectivas de negocio que los componen. De hecho, la multidimensionalidad, que consiste en presentar la información de manera matricial, es una de las características más importantes de estas herramientas. Esta característica distingue la capacidad de estas herramientas respecto a otros sistemas de visualización de datos.

Una vista OLAP permite a los usuarios investigar principalmente la pregunta de: ¿por qué paso? y, como consecuencia, entender qué factores son los determinantes ante un resultado de negocio.

Por ejemplo, una disminución de los beneficios puede ser resultado de un incremento de los costes de producción o una reducción de las ventas. El uso de estas herramientas nos permite investigar este suceso hasta encontrar la causa del mismo.

El proceso completo que se debe seguir para un buen análisis OLAP contempla:

- Proporcionar una vista multidimensional de un almacén de datos relacionales.
- Asegurar el rendimiento acelerado de la herramienta (por ejemplo, almacenando en memoria del servidor el cubo OLAP).
- Llevar a cabo un análisis OLAP, formulando consultas sobre diferentes atributos o ejes.
- Exploración de datos más avanzada, por ejemplo añadiendo más medidas, formateo condicional y cálculos avanzados.
- Exportación de los datos a otro formato (XLS, PDF, CSV, etc.) para su presentación a terceros.

3.3. Cuadros de mando

Tanto los informes como las herramientas OLAP proporcionan una gran cantidad de información a los usuarios, que pueden hacerlas inadecuadas para tomar decisiones rápidas. Además, dichas herramientas responden a preguntas vinculadas al pasado: «¿qué ha pasado?» y «¿por qué ha pasado?», pero no a preguntas vinculadas con el presente o a previsiones de futuro.

Para poder responder a la pregunta «¿qué está pasando?», se necesita otra estrategia/herramienta: el cuadro de mando, que por su funcionalidad se categoriza como una herramienta de monitorización.

Los cuadros de mando provienen del concepto francés *tableau du bord*, y permiten mostrar información consolidada a alto nivel. Se caracterizan por el uso de elementos visuales (gráficos, tablas, alertas, etc.), por presentar una cantidad reducida de aspectos de negocio y por incluir elementos interactivos para potenciar y facilitar el análisis de la información en profundidad y su comprensión.

El cuadro de mando puede estar orientado a información en el pasado o casi en tiempo real, y a toda la organización o a solo un departamento o proceso de negocio. En el primer caso, la información debe extraerse del propio *data warehouse*. Para los cuadros en tiempo real, utilizaremos los sistemas operacionales.

Se debe tener en cuenta que las respuestas obtenidas son el resultado de consultas predefinidas sobre la factoría de información. Normalmente, estas consultas son creadas por un programador de aplicaciones después de haber recopilado las necesidades de cliente. El usuario (cliente) de la herramienta EIS no podrá modificar en ningún momento los parámetros o datos que configuran cada uno de los informes generados. Por lo que, en el caso de necesitar otro informe o una modificación de los existentes, deberá recurrir al programador.

Actualmente, todas las soluciones del mercado incluyen cuadros de mando porque permiten entender muy rápidamente la situación del negocio y son muy atractivos visualmente. Respecto a estos, existen diferencias en un ámbito de madurez del producto, capacidades de visualización, conectividad con diferentes fuentes de datos, versatilidad de interacciones y despliegue multi-dispositivo.

a) Cuadro de mando integral (*balanced scorecard*)

Aparte de comprender el rendimiento pasado y actual de una organización y las razones que hay tras su comportamiento, las organizaciones también necesitan desarrollar sus estrategias de negocio.

Si bien es cierto que toda organización tiene como objetivo a largo plazo su sostenibilidad y la generación de beneficios, cada una de ellas sigue diferentes caminos vinculados a su misión, visión y valores.

La dirección necesita un procedimiento formal para alinear las acciones estratégicas, tácticas y operacionales y poder analizar su evolución. La respuesta a esta necesidad es el cuadro de mando integral.

Muestra de una mirada la comprensión del global de las condiciones del negocio mediante métricas e indicadores clave de desempeño (KPI).

El cuadro de mando integral o *balanced scorecard* es una herramienta que permite analizar una organización respecto a cuatro perspectivas: financiera, de cliente, interna y de innovación y aprendizaje. Esta herramienta surgió en los años noventa como respuesta a la necesidad de analizar las organizaciones respecto al punto de vista financiero, que estaba quedando obsoleto.

El hecho de usar los datos contenidos en un sistema de almacén de datos es lo que realmente lo transforma, y pasa de ser un ejercicio puntual de revisión de la estrategia a una herramienta de alto valor para las organizaciones. Proporciona diferentes beneficios, como:

- Definir y clarificar la estrategia.
- Habilitar una mejor comunicación de la estrategia.
- Alinear objetivos personales y departamentales.
- Vincular el corto y largo plazo.
- Al ser un sistema de control por excepción, permite detectar desviaciones respecto el plan estratégico.

b) Cuadro de mando analítico

Son la base de los sistemas de información para ejecutivos (*executive information system*, EIS) y permiten la visualización rápida y ágil del estado de una determinada situación empresarial, lo que posibilita la detección de anomalías y oportunidades.

Se elabora a partir de los *data mart*, de informes resumen e indicadores clave para la gestión (KPI) que permitan a los gestores de la empresa analizar los resultados de la misma de manera rápida y eficaz. En la práctica, es una herramienta de *query* orientada a la obtención y presentación de indicadores para la dirección (frente a la obtención de informes y listados).

3.4. Otros

Aparte de los formatos de presentaciones vistos anteriormente, existen otros sistemas que ofrecen apoyo a la presentación de la información.

3.4.1. Sistemas de apoyo a la toma de decisiones (*decision support systems*, DSS)

La figura 7 resume los usuarios, el foco y el tipo de información de estos sistemas:

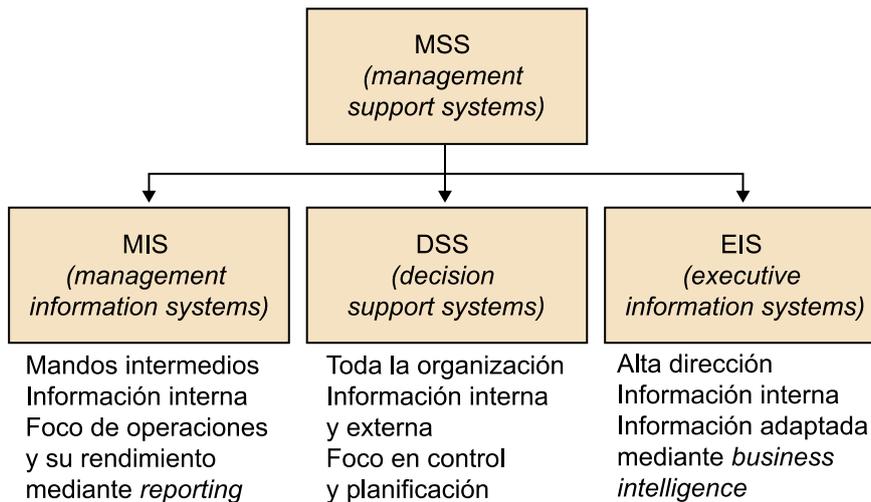


Figura 7. Estructura de sistemas de apoyo a la toma de decisiones
(Fuente: UOC. *Uso de la Factoría de Información Corporativa*)

Si bien la discusión de las diferencias entre estos sistemas sigue siendo tema de investigación, frecuentemente se considera que EIS es un tipo de DSS.

3.4.2. Mapas

En algunos casos, es necesaria la representación sobre el territorio de la información obtenida mediante herramientas de *business intelligence*. En ese caso, las herramientas que sirven para representar en ese formato la información son las herramientas GIS o *geographic information systems*, basadas en mapas. Estas herramientas añaden una capa de visualización sobre la que representan los valores que obtenemos de las herramientas BI.

Podemos afirmar que un mapa es mejor que miles de números. La información geográfica es una parte esencial de muchos de los datos de nuestra organización. Las áreas geográficas tienen límites, nombres y otra información que permite su localización sobre el terreno y relacionar con ello la información propia.

Los mapas son las herramientas más eficientes para visualizar los patrones espaciales. Cuando están cuidadosamente diseñados y presentados, son más que simples elementos decorativos en una presentación estadística, y pueden ayudar a las personas a identificar y resaltar las distribuciones y patrones que puedan no ser evidentes en tablas y gráficos.

Si «una imagen vale más que mil palabras», entonces, «un mapa vale más que mil números». En nuestra era visual, los mapas son una potente forma de informar. Sirven como valiosas herramientas para que los expertos y público en general puedan tomar decisiones, así como para satisfacer una creciente demanda de información en todos los sectores de la sociedad.

Existen muchos tipos de mapas y muchas formas de clasificarlos. Una clasificación especialmente relevante es la que divide a estos en dos grupos cartográficos principales en función del tipo de información que aporten: cartografía base, también denominada fundamental o topográfica, y cartografía temática.

- La **cartografía base** representa el tipo de mapa que originalmente era el objeto principal de la cartografía. Requiere medidas precisas y se basa fundamentalmente en el trabajo de la topografía para obtener la información necesaria que posteriormente se plasma sobre el mapa.
- La **cartografía temática** se centra en la representación de un tema concreto (una variable espacial dada), y puede esta ser de cualquier índole: física, social, política, cultural, económica, etc. Se excluyen de la lista de esos temas posibles los puramente topográficos, que constituyen el objeto de la cartografía base.

Una forma habitual de trabajar es crear una o varias capas temáticas propias con los datos de nuestra organización, que podremos ofrecer a los usuarios sobre cartografías base propias o disponibles de manera pública.

3.4.3. Minería de datos (*data mining*)

La minería de datos consiste en extraer información de alto valor añadido (patrones ocultos, tendencias y correlaciones) a partir de datos en bruto. Lo que se pretende es descubrir información estratégica para construir posteriormente modelos predictivos sobre esta información, modelos que provengan de los productos de la empresa, de sus procesos, de sus clientes, de su competencia, etc., y así poder predecir valores y tendencias en su comportamiento. El concepto de minería de datos no es nuevo, pero la aparición de herramientas potentes, fáciles de utilizar y al alcance de todo tipo de usuarios son elementos recientes que han contribuido a democratizar y potenciar su uso.

Son auténticas herramientas de extracción de conocimiento útil, a partir de la información contenida en las bases de datos. El *data mining* incorpora la utilización de tecnologías basadas en redes neuronales, árboles de decisión, reglas de inducción, análisis de series temporales y visualización de datos.

Es fundamental entender que *data mining* es un proceso; no es simplemente ejecutar un determinado algoritmo que hace alguna tarea, como por ejemplo, una regresión lineal o una serie de cálculos y ya está. Esto no es *data mining*, aunque sí se puede llegar a entender como análisis de datos.

El proceso de análisis de datos suele ser una tarea para matemáticos y estadísticos, pero hay herramientas que facilitan este trabajo a usuarios de negocios o analistas. Aparece un nuevo perfil profesional, denominado *data scientist*.

La minería de datos es un concepto de explotación de datos diferente a los anteriores presentados. Para empezar, no se basa en coeficientes de gestión o información agregada, sino que considera la información en detalle contenida en el almacén de datos. De manera adicional, se persigue identificar una relación o patrón entre los datos, que tenga repercusiones en el negocio más allá de una simple presentación de información.

Por ejemplo: como enfoque empresarial, la minería de datos permite abordar múltiples problemas de una organización: desde la segmentación de clientes hasta la optimización de la cadena de suministro.

Actualmente, los proveedores de soluciones de minería de datos no se están aproximando al mercado con una plataforma genérica, sino con soluciones específicas que resuelven un problema de un área y proceso determinado, que se integran en procesos de negocio y que pueden personalizarse a las necesidades de una empresa particular. Estas soluciones buscan reducir el tiempo de despliegue del proyecto, al reducir la personalización en un 50-60 %. Este enfoque se conoce como analítica de negocio. Es necesario comentar que en el ámbito de los negocios, no basta con identificar patrones ocultos en la información si después la organización no puede actuar de manera ágil con esta información y accionar este conocimiento. El hecho está en que aquellas organizaciones que son realmente capaces de ejecutar una estrategia de analítica de negocio correctamente están generando nuevas ventajas competitivas.

3.4.4. Autoservicio BI

Podemos agrupar las diferentes tendencias actuales bajo el paraguas que podríamos denominar *selfservice BI*. Encontramos múltiples opciones: *data discovery*, *visual analytics*, *autodiscovery*, *query by exemple*, *discovery analysis*, *business discovery*, *natural analytics*, etc.

Las características más relevantes de estas herramientas son:

- Soportan técnicas de **análisis visual** que facilitan la comprensión rápida de los datos y permiten, asimismo, llevar a cabo presentaciones claras y eficaces que ayuden en la toma de decisiones.
- Facilitan el análisis de **grandes cantidades** de datos.
- A menudo, centradas en el **business discovery**: descubrimiento de errores u oportunidades de mejora.
- Utilizan la **lógica asociativa** (AQL), técnica que efectúa los análisis y cálculos en memoria, obteniendo con ello un tiempo de respuesta excelente.
- Permiten hacer **análisis interactivo** apoyándose en ágiles funcionalidades de visualización y gestión de datos, lo que hace posible un análisis libre sobre el modelo de datos importado en la herramienta. Es decir, el usuario

tiene la posibilidad de interactuar con los datos: comparar, filtrar, conectar unas variables con otras, etc.

- Se apoyan en una **interfaz intuitiva** que facilita la exploración de datos orientada tanto a perfiles TI como analistas de negocio. La herramienta se adapta al usuario, no al revés.
- Agilidad y rapidez en el manejo de datos, apoyándose en tecnologías ***in-memory***.
- Se adaptan fácilmente a múltiples dispositivos: ordenador, tableta, teléfono inteligente, etc. (en inglés, ***responsive***).

Las principales diferencias entre el autoservicio BI y el BI tradicional son las siguientes:

- Mayor flexibilidad en el análisis, el usuario decide qué usa como dimensión, indicador, jerarquías de análisis, etc. Al no existir una capa semántica, el usuario se enfrenta directamente a la BDD, lo cual puede ser arduo cuando no hay mucha documentación y dar lugar a la creación de vistas resumen (típico tablón) con campos autoexplicativos partiendo de las tablas origen.
- No es necesario crear repositorios de datos tipo *data warehouse*. Acceso directo a fuentes sin procesos ETL previos. Esto da flexibilidad, pero podemos encontrarnos datos en bruto poco validados.
- Mayor independencia respecto al área de IT, pero menor control sobre los entornos de análisis. Libera a IT de trabajos recurrentes de poco valor.
- Más apropiado para organizaciones pequeñas, con un volumen y complejidad de información menores. Ciclos cortos de desarrollo y costes más bajos (considerando costes productos + consultoría, compite con soluciones BI *open source*)
- Ciclos más cortos en procesos de análisis *ad hoc*.

3.4.5. Sistemas de búsqueda empresarial en lenguaje natural

A medida que las organizaciones generan más y más datos e información, ha crecido la necesidad de capturar, extraer, consolidar y acceder a la información de manera rápida y eficiente.

Una posible solución al acceso de la información la proporcionan los sistemas de búsqueda empresarial. Estos sistemas tienen el objetivo de indexar y hacer accesibles todos los contenidos empresariales de múltiples fuentes, incluyendo ficheros, bases de datos, intranets, sistemas de gestión documental, sistemas operacionales, EIS, *business intelligence*, correo electrónico y bases de datos.

Es decir, los usuarios pueden preguntarle a la plataforma y esta responde a las peticiones con gráficos: una mezcla de análisis y visualización de datos con inteligencia artificial. Este es un paso importante dentro del mundo del BI, en la carrera por facilitar el uso de los datos.

Estos sistemas han evolucionado no solo para proporcionar acceso a contenidos específicos mediante lenguaje natural, sino también para correlacionar información de distintas fuentes de información y usar esta correlación en los resultados de las consultas.

Por ejemplo, la búsqueda del nombre de un cliente puede devolver la información del perfil de cliente (del CRM), el contrato del servicio que la empresa tiene con él (del gestor documental) y la información histórica de sus compras (del ERP o la FIC).

3.4.6. *Big data*

La visualización de datos masivos se tratará en otra asignatura específica sobre este tema.

3.4.7. *Webhousing y mobile BI*

Desde hace unos años, la evolución de la tecnología ha estado centrada en internet. Con la llegada y la democratización de las tecnologías móviles, cualquier nueva tecnología busca estar presente y vinculada tanto a internet como a la movilidad.

- El concepto de *webhousing* nace de la confluencia de los almacenes de datos con Internet. Esta simbiosis crea un nuevo esquema de información en el que los clientes tienen a su disposición grandes cantidades de información.
- Una de las tendencias actuales es ofrecer el almacén de datos como un servicio (como por ejemplo, Amazon Redshift), caso en el que el acceso a los datos siempre sería por internet. Igualmente, existe la posibilidad de no solo ofrecer el *data warehouse* como un servicio, sino también la capa de acceso y análisis como servicios, lo que se conoce en inglés como *BI as a service*. En estos casos, las mismas arquitecturas descritas son válidas, pero es el proveedor quien las soporta mediante *cloud computing*.
- El concepto de *mobile BI* hace referencia al acceso de los datos a través de dispositivos móviles, encapsulando estos accesos en aplicaciones específicas de valor añadido, como por ejemplo un comparador de precios.

Aunque cada vez el usuario dispone de mayor flexibilidad, el desarrollo de nuevas técnicas de visualización o representación y la existencia de más sitios web interactivos pueden causar problemas para las organizaciones. Cada vez es más fácil para los usuarios, ya sea por accidente o a propósito, distorsionar o tergiversar los datos y luego hacer que estas distorsiones e interpretaciones erróneas estén ampliamente disponibles para otras personas.

Por tanto, es necesario que las organizaciones de estadística tengan una política clara sobre cómo aplicar y ofrecer nuevas técnicas de visualización o representación.

4. Herramientas de apoyo a la presentación de datos

Se pueden mostrar los datos de muchas maneras diferentes, desde sencillos gráficos de barras a diagramas de dispersión más complejos, mapas temáticos y pirámides de población animadas. No faltan ayudas técnicas: libros escritos en torno a la visualización de los datos; apuntes en páginas web dedicados al tema; y una amplia gama de software y programas descargables, disponibles para cualquier finalidad.

4.1. Metodologías

Son distintos los intentos de establecer métodos capaces de lidiar con los nuevos formatos de datos y conjuntos de datos. Permitiendo que formatos de datos específicos, como espectrales o espaciotemporales, puedan visualizarse adecuadamente; consiguiendo adaptar la tecnología para hacer frente a datos heterogéneos (y de diferente credibilidad); y haciendo posible enfatizar los aspectos que son relevantes para la visualización.

1) Metodología Lengler y Eppler:

Pretende definir y compilar métodos de visualización existentes, con el fin de desarrollar una revisión sistemática basada en la lógica y el uso de la tabla periódica de elementos, revisada con 100 métodos de visualización y una propuesta de cómo usar cada uno de ellos.

A PERIODIC TABLE OF VISUALIZATION METHODS

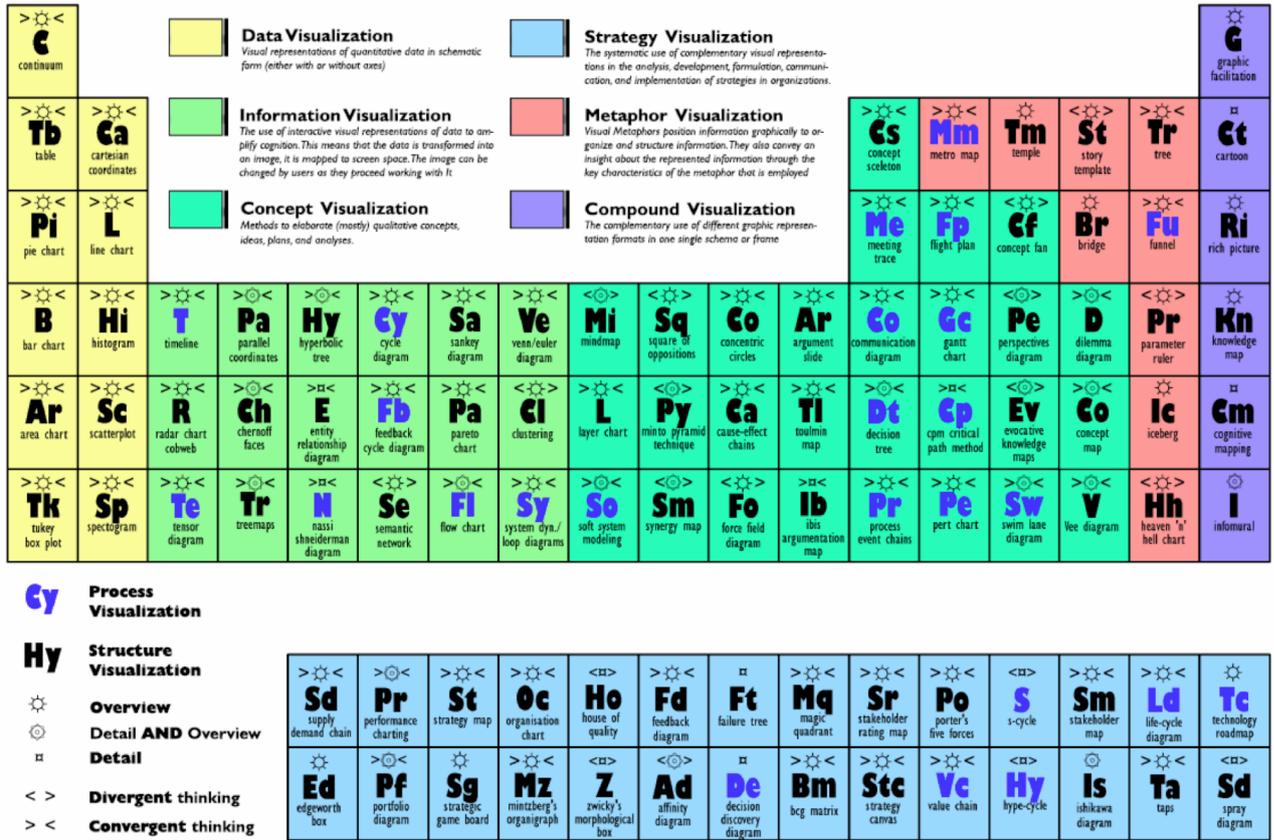


Figura 8. Tabla periódica de métodos de visualización de datos (Fuente: http://www.visual-literacy.org/periodic_table/periodic_table.pdf)

2) Metodología Lee Wittschen:

Para que BI apoye la toma de decisiones, los datos deben ser los correctos e interpretados adecuadamente.

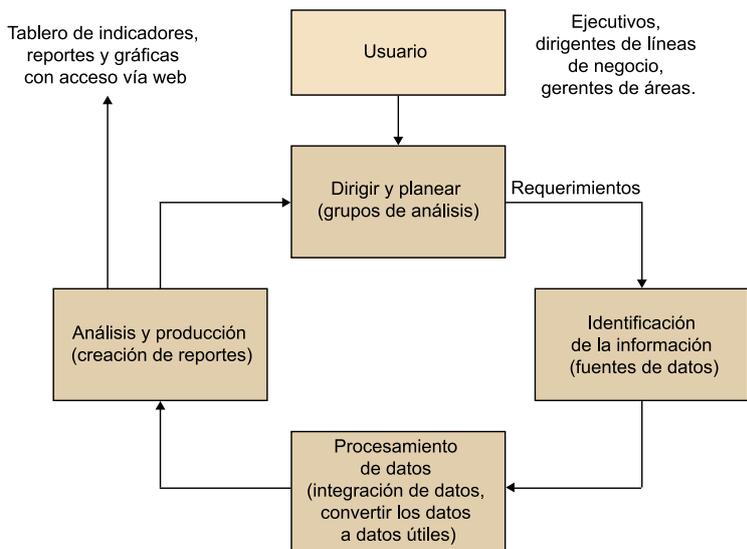


Figura 9. Proceso Lee Wittschen

- Fase 1. Dirigir y planear

Mediante un grupo de analistas, se definen los requerimientos para la gestión de la empresa y se formulan preguntas e hipótesis para alcanzar el objetivo y cubrir las necesidades.

- Fase 2. Recolección de información
Con la finalidad de obtener los resultados esperados, se deben identificar las bases de datos donde reside la información, sin importar que provenga de sistema externos.
- Fase 3. Procesamientos de datos
Se deben integrar los datos en un mismo formato para que puedan ser analizados. Los datos de la etapa 3 son interpretados y procesados.
- Fase 4. Análisis y producción
Se crean reportes personalizados, que trazarán los cursos de acción que hay que seguir, y proporciona datos o indicadores relevantes para la toma de decisiones.
- Fase 5. Difusión
Se generan los reportes personalizados y esperados, para su interpretación, y se trazan y gestionan los cursos de acción.

3) Metodología Ann K. Emery:

La visualización de datos mezcla tecnología, ciencia y creatividad a partes iguales. La personalización es el denominador común, aunque no exime de la observación de algunas reglas, como las propuestas, que contribuyen a mejorar los resultados. Para una evaluación de la calidad de la visualización de datos en mayor profundidad, se puede acceder al *checklist* que proponen Ann K. Emery y Stephanie Evergreen, en el que se toman en consideración aspectos como:

La forma de presentación del texto, en los casos en que sea necesario incluirlo: deberá ser legible, concreto y preciso, preferiblemente en disposición horizontal y susceptible de aplicarse directamente a los elementos que componen el gráfico o diagrama.

- La disposición gráfica debe ser clara, de proporciones adecuadas, con un mínimo dos dimensiones y siempre aplicando criterios de orden.
- El uso del color: debe tener el suficiente contraste, no perder su capacidad de informar ni siquiera en el caso de una impresión en blanco y negro, y debe prever las limitaciones de usuarios con problemas de diferenciación de colores.
- Presencia de líneas: los lugares donde están presentes y su apariencia, que nunca debe distraer la atención del contenido presentado.

- **Visión global:** que tiene que permitir el énfasis en los detalles más relevantes, la adecuación del tipo de gráfico a los datos presentados, la contextualización y un nivel de precisión óptimo.

4.2. Técnicas y componentes

Las herramientas y técnicas emergentes están proporcionando nuevas oportunidades para la visualización de datos y para hacerlos más interesantes a los usuarios.

Los generadores de tablas dinámicas, gráficos y mapas permiten a los usuarios manipular los datos y crear sus propias representaciones. La animación y el vídeo son formatos atractivos, similares a la televisión. Hacen un buen trabajo ilustrando los cambios a lo largo del tiempo, e incluyen descripciones verbales o textuales que explican el significado detrás de los números. También están surgiendo nuevos tipos de visualizaciones, como *sparklines* y nubes de etiquetas, que proporcionan alternativas para ilustrar la información.

Sparklines

Son pequeños gráficos de líneas del tamaño de palabras, que muestran tendencias en el tiempo.

No todos los tipos de gráficos son adecuados para un conjunto concreto de datos. Algunos de ellos solo valen para un fin, y otros se adaptan a varias clases de datos. Para tomar la decisión de cuál usar, debemos tener en cuenta, por un lado, el tipo de medida usada y, por otro, las características del conjunto de datos: si son series temporales o no, si interviene una o distintas variables.

Si lo que deseamos es reflejar el **tipo de medida** que estamos usando, elegiremos el tipo de gráfico según esta figura:

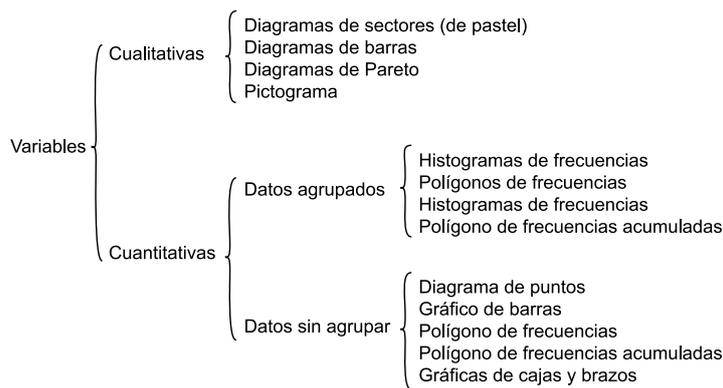


Figura 10. Tipos de gráfico según tipo de medida

También podemos usar esta «receta rápida» de aplicación general (aunque siempre con excepciones, ya que la representación y simbolización contiene, no lo olvidemos, elementos subjetivos), según estos puntos:

- Para las **variables cualitativas** se emplean las variables visuales color, forma y textura, en la medida en que sea posible, según el tipo de objeto geométrico que hay que simbolizar.
- Para las **variables cuantitativas**, el valor del color y el tamaño son las más adecuadas, y esta última es la única que permite transmitir toda la información en el caso de variables de tipo razones. El tono de color puede emplearse, pero se debe elegir una gama de tonos que presente algún tipo de lógica que permita establecer un orden.

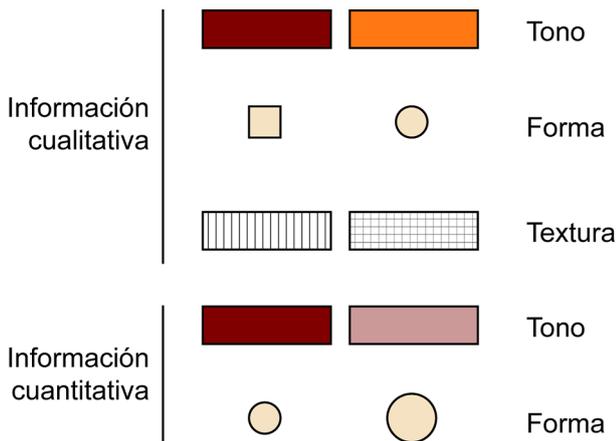


Figura 11. Forma, color y textura según el tipo de variable

La **animación** y el vídeo son dos importantes técnicas emergentes de visualización de datos. Cuando se tiene en cuenta la popularidad de la televisión y el cine, no es de extrañar que a los usuarios les guste la idea de recibir mensajes a través de imágenes en movimiento. Este formato hace más fácil contar la historia, mediante la combinación de descripciones en audio o texto, con ilustraciones gráficas que explican el significado que se esconde tras los números.

Un buen ejemplo del uso de animaciones para explicar estadísticas es la pirámide de población del Reino Unido, 1971-2081

http://www.statistics.gov.uk/populationestimates/svg_pyramid/uk/index.html

Las pirámides dinámicas de población, desarrolladas por varias organizaciones de estadística, incluidas la Oficina de Estadísticas Nacionales del Reino Unido y Estadísticas de Canadá, son buenos ejemplos de la combinación de animación con interactividad en una interfaz sencilla. Los usuarios pueden hacer clic para ver cómo la forma de la pirámide de población cambia con el paso del tiempo. También pueden interactuar con el gráfico seleccionando grupos de edad y fijándose con detalle en los números y en las proporciones en la población total.

Hans Rosling, cofundador de Gapminder, ha tenido un gran éxito con el uso de la animación para ilustrar datos. Rosling ha alcanzado una audiencia masiva a través de pequeños vídeos en línea (*gapcasts*), los cuales son conferencias sobre distintos temas, como la mortalidad materna, la globalización, la energía, etc. (<http://www.gapminder.org/video/gap-cast/>).

Las *sparklines* tienen la ventaja de mostrar una gran cantidad de información de un vistazo, y se pueden colocar junto a palabras que expliquen su significado. Tufte (2006) fue el primero en proponerlas. Estos «directos y sencillos gráficos del tamaño de palabras» mejoran la presentación de los datos con una representación visual, sin ocupar mucho espacio.

La siguiente figura muestra un ejemplo de *sparklines* utilizadas para ilustrar las fluctuaciones de las tareas efectuadas por los miembros de un equipo de proyectos.

Team Member	Total Tasks Completed	w1	w2	w3
Julie	 46%	13	15	19
John	 45%	11	18	11
Jabba the hut	 -20%	15	14	14
Johnson	 6%	18	17	14
Jeremy	 43%	14	20	10
Josh	 -33%	15	12	19

Figura 12. *Sparklines* elaboradas en MS Excel

Una **nube de etiquetas** (a veces también denominada nube de palabras) es una representación visual de la frecuencia de palabras o etiquetas en un texto o *dataset* concreto. Se puede ver a menudo en páginas web como una lista de categorías, donde cada palabra es un vínculo que conduce al usuario a más información relacionada con esta palabra.

Las nubes de etiquetas son una forma útil de identificar los términos comunes de un texto y construir taxonomías de palabras clave.

Ejemplo

El siguiente ejemplo ha sido creado utilizando el texto de este módulo, e ilustra claramente las palabras clave.



Figura 13. Nube de palabras

4.3. Recursos en línea

Ya sabemos que la imagen es un elemento muy poderoso. El 90 % de la información que se transmite al cerebro es visual, y esta se procesa 60.000 veces más rápido que el texto, por lo que es lógico que acompañemos nuestros datos con imágenes o recursos gráficos apropiados.

Existen multitud de recursos disponibles:

a) Guías de cromatismo, estilos y tipos de gráficos:

- *Color advice for cartography* (<http://colorbrewer2.org>)
- Tipos de gráficos, ¿cuál uso? (http://www.ine.es/explica/explica_pasos_tipos_graficos.htm)

IN e Instituto Nacional de Estadística

Explica
te ayudamos con la Estadística

Inicio / Primeros pasos / **Tipos de gráficos ¿cuál uso?**

◀ 1 / 28 ▶

¿Qué es Explica?

Primeros pasos

Estadísticas oficiales

Estadística y mucho más

Juega con nosotros

Un poco de historia

Olimpiada Estadística

Índice

- 1- Introducción
- 2- Tipos de datos
- 3- Gráfico de barras
- 4- Pirámide de población
- 5- Gráfico de líneas
- 6- Gráfico de Pareto
- 7- Gráfico de sectores
- 8- Pictograma
- 9- Gráfico de dispersión
- 10- Cartograma
- 11- Bibliografía

[Continúa](#)

© INE 2016 www.ine.es Accesibilidad Aviso legal Paseo de la Castellana, 183 - 28071 - Madrid - España

- *Google Chart types* (<https://developers.google.com/chart/>)

Google Charts

HOME GUIDES REFERENCE SUPPORT

Overview

Hello, Charts!
Quickstart
Load the Charts Library
Prepare the Data
Customize the Chart
Draw the Chart
Draw Multiple Charts

Chart Types

- Chart Gallery
- Annotation Charts
- Area Charts
- Bar Charts
- Bubble Charts
- Calendar Charts
- Candlestick Charts
- Column Charts
- Combo Charts
- Diff Charts
- Donut Charts
- Gantt Charts
- Gauge Charts
- GeoCharts
- Histograms
- Intervals
- Line Charts
- Maps
- Org Charts
- Pie Charts

Geo Chart

Scatter Chart

Column Chart

Histogram

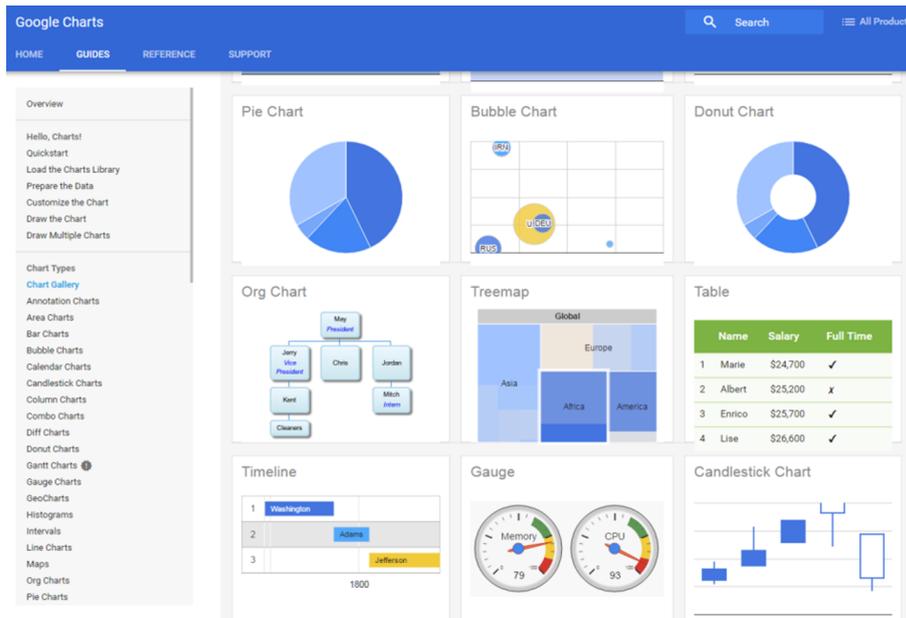
Bar Chart

Combo Chart

Area Chart

Stepped Area Chart

Line Chart



- *Pentaho Visualization plugins examples* (<http://wiki.pentaho.com/display/COM/Visualization+Plugins>)

Visualize Your Data in a Whole New Light

12 Ways to Examine Your Data Like Never Before

We have unveiled 12 useful visualizations that can be downloaded and plugged into the Pentaho platform - enabling users to make decisions even faster. If you are interested in learning how to create your own visualization plugins, please visit the [Pentaho InfoCenter](#) for details.

Visualizations



b) Bancos de imágenes y recursos gráficos:

Recopilamos a continuación varios sitios web en los que se pueden encontrar recursos gráficos para hacer infografías o ilustraciones.

- Free Digital Photos (<http://www.freedigitalphotos.net/>)
- Freemages (<http://www.freemages.com/>)
- Morgue File (<http://www.morguefile.com/>)

- Icojam (<http://www.icojam.com/>)
- Freepick (<http://www.freepik.es/>)
- Flaticon (<http://www.flaticon.com/>)

c) Bancos de **tipografías**:

- Dafont (<http://www.dafont.com/es/>)
- Google Fonts (<https://www.google.com/fonts>)

d) **Librerías** gráficas y de componentes:

- *D3: building reusable charts* (<http://misoproject.com/d3-chart/>)
- *Mondrian: general purpose statistical data-visualization* (<http://www.theusrus.de/Mondrian/>)

e) **Comunidades** en torno a los datos:

Existen sitios web que están añadiendo una nueva dimensión a las presentaciones visuales, a través de la construcción de comunidades en línea sobre la visualización y el intercambio de datos. Estos sitios permiten a los usuarios cargar conjuntos de datos y crear gráficos, para su intercambio y discusión con otros usuarios. Otras aplicaciones, denominadas aplicaciones web híbridas (*mashups*), combinan datos o funcionalidades de dos o más fuentes para crear un servicio nuevo.

Por ejemplo, la UNECE ha subido cuatro bases de datos a Swivel, con resultados distintos. En el lado positivo, cada una de estas *datasets* ha sido vista más de 5 000 veces durante los primeros 18 meses, lo que indica que los datos han llegado a muchos usuarios. <http://www.swivel.com/users/show/1005968>

Aunque hasta ahora el éxito ha sido desigual, este tipo de comunidades en línea sin duda proporcionan una forma relativamente fácil de llegar a más usuarios y es, por lo tanto, un área emergente en la visualización de datos que hay que tener en cuenta.

4.4. Herramientas de visualización

En la actualidad, existe una gran variedad de herramientas analíticas y de BI, lo cual puede dificultar la elección de la solución más adecuada, que en cualquier caso tendrá que ver con el tipo de empresa, tamaño, sector, objetivos y procesos que se quieran cubrir.

Los «cuadrantes mágicos» son una herramienta creada y promovida por la empresa Gartner, consistente en una representación gráfica del mercado analizado en un momento determinado.

De acuerdo con lo definido por Gartner, los líderes son aquellos fabricantes que tienen una visión clara de la dirección del mercado y que desarrollan competencias para mantener su posición actual.

Estos cuadrantes ayudan a las organizaciones a identificar y poder diferenciar a los proveedores de servicios en el campo BI y plataformas analíticas. Deben ser considerados como una herramienta, y no como una guía.



No es fácil encontrar información clara sobre tarifas de licenciamiento y soporte de software comercial de *business intelligence*. En este enlace, encontramos un estudio que nos puede orientar:

<http://www.slideshare.net/oktopuslu/bi-comparison-of-open-source-and-traditional-vendor-4327259>

No siempre se han considerado en las comparativas de herramientas las opciones de BI código abierto (*open source*) que existen en el mercado, frente a las opciones BI propietario.

- Las primeras, sin coste de adquisición, pero con una cantidad importante de horas de desarrollo.
- Y las segundas, con el valor añadido de un apoyo técnico muy interesante, que en las *open source* se ofrece a través de las comunidades de usuarios que han adoptado estas herramientas.

Teniendo todo esto en cuenta, elegir entre herramientas de visualización de datos reduce su complicación, puesto que la meta está clara: ya se sabe lo que se necesita.

Es el momento de evaluar otras variables en función de los siguientes criterios:

1) Criterios de negocio:

- Velocidad de implementación
- Escalabilidad
- Precio
- Coste de licencias / apoyo técnico
- Viabilidad a largo plazo

2) Criterios técnicos:

- Orientación a usuarios finales
- Capacidades analíticas
- Visualización interactiva
- Capacidades de modelado
- *Drill-down* visual
- Controles visuales
- Soporte de escritorio
- Entorno de desarrollo
- Integración con otras herramientas/plataformas/sistemas
- Conjuntos de datos *in memory*

En este ejemplo, vemos algunos criterios y factores que podemos tomar en consideración cuando evaluamos varias tecnologías, antes de su elección:

- ¿Ha tenido un crecimiento muy espectacular en los últimos años?
- ¿Usa lógica asociativa (AQL), haciendo los análisis y cálculos en memoria?
- ¿Permite la elaboración de prototipos rápidos?
- ¿Soporta grandes volúmenes de datos (*big data*)?

- ¿Forma parte de un paquete ofimático BI completo, o es una solución específica?
- ¿La relación precio-soporte es interesante?
- En los últimos años, ¿ha ido escalando posiciones en los cuadrantes de Gartner?
- ¿Cubre las funcionalidades requeridas?
- ¿Existen otras organizaciones que usan esa herramienta?
- ¿Cómo gestiona los metadatos?
- ¿Dispone de un buen sistema de versionado y control del código fuente?
- ¿Qué política de nuevas versiones, mejoras y resoluciones de *bugs* aplica?
- ¿Es accesible y dispone de visualización desde iPad, Android u otros?
- ¿Dispone de un sistema que permita hacer *write-back* para establecer presupuestación, *forecasts*, simulación o reglas de negocio?
- ¿Dispone de un motor OLAP?
- ¿Dispone de herramientas de *data mining*?
- ¿Soporta diferentes arquitecturas y entornos como Windows o Linux?
- ¿Puede ser implementada y usada en semanas, dependiendo de la complejidad y volúmenes de datos?

4.5. Herramientas de apoyo

Encontramos distintas herramientas que complementan las funcionalidades de las herramientas del apartado anterior, y que cubren tareas técnicas de optimización o refinamiento para su buen funcionamiento.

- **Optimizador de consultas (*query analyzer*)**
Es una herramienta de optimización de consultas. Con ella se pueden visualizar y solucionar problemas de los registros de consulta generados al ejecutar informes de consulta dinámica. Algunas bases de datos incorporan un planificador de consultas que lleva a cabo funciones equivalentes
- **Agregador de datos**
El asesor de agregación analiza cubos dinámicos para recomendar agregados que mejoren el rendimiento del cubo. El asesor de agregación puede aplicar los agregados en base de datos automáticamente a las tablas de base de datos y al modelo. Puede decidir si desea aplicar los agregados en base de datos inmediatamente, o planificarlos para más adelante. Cuando aplica recomendaciones en base de datos, el asesor de agregación puede planificar un evento para cargar datos más tarde.
- **Integración/conectividad**
Necesitamos ineludiblemente manejar múltiples fuentes y formatos de datos en las bases de datos. La mayoría de las herramientas aceptan formatos como Excel, Access y texto; pueden acceder a muchas bases de datos como Microsoft SQL Server, MySQL, Oracle o Greenplum; y también tienen la

posibilidad de usar las API (*application programming interface*) de otras herramientas, para la extracción sistemática de datos; conjuntos de datos de Hadoop, datos de *streaming* y servicios en la nube.

- **Portales de distribución, colaboración y alertas**

Estos sistemas mandan alertas y envían correos electrónicos de manera automática para notificar los eventos a los responsables de la toma de decisiones de su organización a medida que tienen lugar, para que puedan tomar las decisiones de forma rápida y eficaz. También se pueden ver y abrir espacios de trabajo donde compartir y manipular datos y/o análisis. En ellos podremos utilizar comentarios, actividades y software social para la toma de decisiones colaborativa.

Glosario

almacén de datos *m* Bases de datos orientadas a áreas de interés de la empresa que integran datos de distintas fuentes con información histórica y no volátil, y que tienen como objetivo principal hacer de apoyo en la toma de decisiones. Puede ser corporativo o departamental.
en data warehouse

análisis interactivo *m* Análisis libre sobre el modelo de datos, apoyándose en ágiles funcionalidades de visualización y gestión de datos, lo que permite interactuar con los datos.

análisis visual *m* Técnicas que facilitan la comprensión rápida de los datos y permiten, asimismo, llevar a cabo presentaciones claras y eficaces que ayuden en la toma de decisiones.

BI *m* Véase **business intelligence**.

BI estratégico *m* BI que llevan a cabo aquellas organizaciones que están madurando en su estrategia de inteligencia de negocio y se están apalancando en el BI para empujar su transformación estratégica.

BI operacional *m* aplicado sobre un *data mart* operacional concreto.

BI táctico *m* Despliegue de una estrategia de *business intelligence* que dé respuesta a las necesidades de información para la toma de decisiones de los mandos medios de una organización.

big data *m* Conjunto de estrategias, tecnologías y sistemas para el almacenamiento, procesamiento, análisis y visualización de conjuntos de datos complejos.

business discovery *m* Descubrimiento de errores u oportunidades de mejora, soportado en herramientas específicas.

business intelligence *m* Habilidad de transformar los datos en información y la información en conocimiento, y así optimizar el proceso de toma de decisiones de la dirección de una empresa.

sigla: **BI**

carga cognitiva *f* Cuánto tiene que esforzarse el lector para entender lo que se está tratando de comunicar.

cuadro de mando *m* Formato de presentación que permite mostrar información consolidada a alto nivel, a través del uso de elementos visuales (gráficos, tablas, alertas, etc.) y elementos interactivos.

data mart *m* Subconjunto de los datos del *data warehouse* con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica.

data mining *m* Véase **minería de datos**.

data warehouse *m* Véase **almacén de datos**.

geographic information systems *m pl* Véase **sistema de información geográfica**.

GIS *m* Véase **sistema de información geográfica**.

minería de datos *f* Extracción de información de alto valor añadido (patrones ocultos, tendencias y correlaciones) a partir de datos en bruto.
en data mining

OLAP *m* Siglas que hacen referencia a las herramientas de análisis, normalmente multidimensional, *en on-line analytical processing*.

reporting *m* Soluciones que permiten diseñar y gestionar (distribuir, planificar y administrar) informes.

responsive *m* Propiedad de un elemento que se adapta fácilmente a múltiples dispositivos: ordenador, tableta, teléfono inteligente, etc.

sistema de búsqueda empresarial *m* Herramienta con el objetivo de indexar y hacer accesibles todos los contenidos empresariales de múltiples fuentes.

sistema de información geográfica *m* Herramienta que sirve para representar la información basadas en mapas.

en geographic information systems
sigla **GIS**

WAQR *f* *Web Ad-hoc Query Reporting*.

webhousing *m* Consulta del almacén de datos por web.

Bibliografía

Cano, J. L. (2007). *Business intelligence. Competir con información*. Depósito Legal: M-41185-2007

Emery, A. K. *Data visualization checklist*. <http://annkemery.com/portfolio/dataviz-checklist/>

Lantares Solutions. *Visualización de Datos*.

Lengler, R.; Eppler, M. J. *Towards a Periodic Table of Visualization Methods for Management*. Suiza: Institute of Corporate Communication, University of Lugano, Switzerland

Lowering the Cost of Business Intelligence With Open Source (2010). Technology White Paper.

<http://www.gestiopolis.com/inteligencia-de-negocios/>

Naciones Unidas (2009). *Cómo hacer comprensibles los datos: Una guía para presentar estadísticas*. Ginebra.

