

## **1.6. Explotación de las bases de datos**

Las bases de datos se construyen para contener una gran colección de datos. Cualquier organización genera una cantidad de datos asombrosa, desde las pymes hasta las grandes empresas. En la actualidad, estos datos son la fuente de nuevas herramientas en la explotación de las bases de datos. Estas nuevas herramientas están evolucionando vertiginosamente a la vez que las propias tecnologías lo hacen. Al tener mayor ancho de banda en las conexiones, sistemas de almacenamiento ultra masivos, sistemas de procesamiento de la información de alta velocidad y mayor preparación de los profesionales que tratan con estos, las tecnologías emergentes sobre la explotación de bases de datos se están desarrollando a gran velocidad.

Los almacenes de datos constituyen la arquitectura base para el sistema de soporte a decisiones y la minería de datos. Las bases de datos constituyen el acúmulo de datos que la organización requiere para sus gestiones de cada día, como pueden ser ventas, compras, pedidos, trazabilidad, producción, logística, marketing, gestión de personal, etcétera.

Los **sistemas de soporte a decisiones** no son sistemas de información para el procesamiento de esta, su almacenamiento y flujo, sino que, más bien, son sistemas que, a través de técnicas emergentes de análisis de los datos almacenados, se explotan para la toma de decisiones: qué producto es más rentable, cómo hacer más rentable un producto que no lo es, quién lo adquiere, cómo lo adquiere, qué país lo necesita más, qué país lo compra menos, cómo hacer que se venda más, mejor, con mayor rentabilidad, a más velocidad, con menos agentes intermediarios, etc. Las tomas de decisiones no tienen por qué ser siempre comerciales, pueden ser políticas, sociales, de salud, de prevención de riesgos, de control de amenazas naturales o humanas, y un largo etcétera.

Los **almacenes de datos o Data Warehouse** aportan una metodología para el análisis de datos masivos. A veces, es necesario aglutinar esa ingesta de datos en divisiones más pequeñas con el fin de centrarlos por temas o áreas de negocio, y así se puedan tomar decisiones más localizadas. Este subconjunto de datos de todo el DataWarehouse se denomina **Data Mart**. Se consultan mediante herramientas OLAP (On Line Analytical System) para la construcción de sistemas de información para directivos (EIS) y para los sistemas de ayuda a la toma de decisión (DSS).

En la actualidad, los algoritmos que analizan esos datos usan la inteligencia artificial y los esquemas de los sistemas expertos. Los sistemas expertos son un área de la inteligencia artificial que simulan el proceso de aprendizaje. Estos almacenan datos y conocimiento con el fin de obtener conclusiones y ayudar a tomar decisiones.

Los tipos de **sistemas de información** más utilizados desde un punto de vista empresarial y su función son:

- Los sistemas de procesamiento de transacciones, TPS, Transaction Processing System: almacenar y procesar las transacciones operacionales que se usarán en los sistemas de apoyo a la toma de decisiones.
- Los sistemas de información administrativa, MIS, Management Information System: facilitar la información de interés sobre la situación general del negocio a la gerencia de la organización.
- Los sistemas de soporte a la decisión, DSS, Decision Support System: combinar analizar los datos para resolver problemas.
- Los sistemas de soporte a ejecutivos, EIS, Executive Information System: extraer información de interés para las metas estratégicas.

- Los sistemas para la toma de decisiones en grupo, GDSS, Group Decision Support System: compartir información para las decisiones conjuntas.
- Sistemas expertos de soportes a la toma de decisiones, EDSS, Expert Decision Support Systems: desarrollan conocimiento en áreas concretas y actúan como consultores expertos.

#### 1.6.1. Los sistemas transaccionales

El concepto de transacción se introduce para gestionar la complejidad que conlleva los múltiples accesos al mismo recurso de información. Estos procesos, considerados de ejecución concurrente, pueden acarrear errores sobre la fiabilidad de los datos, mantenimiento y recuperación ante fallos.

Se llama **transacción** a un conjunto de operaciones que se deben hacer todas juntas para que no afecten a su acceso. Esta unidad lógica permite asegurar la integridad de los datos.

El sistema gestor de base de datos debe implementar operaciones que permitan recuperarse de fallos. Además, las operaciones que los usuarios desarrollan no deben interferir entre ellas. El usuario debe acceder a su esquema como si tuviera acceso exclusivo. Esto se logra con técnicas de programación que explotan procesos concurrentes.

#### Sistemas On-Line Transaction Processing, OLTP

Estos sistemas se caracterizan por que sistematizan las transacciones en sus bases de datos y usan sistemas online para la inserción de grandes volúmenes de datos denominados **datos operativos**.

Los **sistemas OLTP** tienen características que hacen que no sean tecnologías óptimas para los sistemas de toma de decisión. Estas son principalmente:

- Los datos proceden de fuentes muy heterogéneas, lo que provoca falta de consistencia.
- Tienden a organizarse por criterios geográficos, ya que su almacenamiento se hace muy rápido y se persigue no provocar cuellos de botella.
- Se prioriza las operaciones de actualización e inserción de datos sobre la resolución de consultas, por lo que la redundancia es casi nula, pero las consultas complejas pueden llegar a ser casi inoperables, pudiendo ocurrir que la velocidad de entrada de datos es mayor que el tiempo de respuesta a los accesos concurrentes de los usuarios..

#### 1.6.2. Sistemas de soporte a la decisión, DSS

Los **sistemas de soporte a la decisión** son una herramienta de la inteligencia de negocio, Business Intelligence, cuyo objetivo principal es el análisis de datos. Supone una herramienta de explotación de bases de datos orientadas al almacenamiento Data Warehouse. corporativo o Data Mart de inmenso tamaño y que pueden resultar minas de información para la toma de decisiones empresariales. Los programas informáticos que los componen usan métodos analíticos como algoritmos de optimización, análisis de decisión, programas de planificación de rutinas, etc., y resuelven los problemas inherentes de los programas de gestión de generaciones anteriores. Esta herramienta acompaña a otras en la inteligencia de negocio, como los cuadros de mando o los

sistemas de información ejecutiva.

Los tipos de sistemas de soporte a decisiones más usados en la actualidad son:

- Sistemas de apoyo a decisiones de grupo, GDSS.
- Sistemas de información ejecutiva, EIS.
- Sistemas de información gerencial, MIS.
- Sistemas expertos basados en inteligencia artificial, SSEE.

### Componentes de un sistema de soporte a la decisión

Los componentes principales de un DSS son:

**Componente de diálogo:** el conjunto de opciones para el usuario para dirigir las acciones del sistema. Está formado por la base de conocimiento, lo que el usuario conoce sobre la decisión y cómo usar el sistema, el lenguaje de acción y de presentación, las propias acciones del usuario y el estilo de diálogo, es decir, las combinaciones de los elementos anteriores.

**Componente de datos:** los datos suponen la entrada para el procesamiento de modelos y lo manejan las fuentes de datos, y los propios almacenes de datos. Estos almacenes se separan para el conjunto de aplicaciones de apoyo a la decisión comúnmente conocido como procesamiento analítico en línea (OLAP).

**Componente modelo:** suministra la capacidad de análisis. Los más usados son los de optimización, los descriptivos, los probabilísticos, los estratégicos, los tácticos y los operacionales.

### El área de ayuda a la toma de decisiones

Dentro de una organización el área de ayuda a la toma de decisiones puede abarcar, a su vez, todas o algunas de las áreas que se muestran a continuación:

El área de procesamiento analítico en línea (Online Analytical Processing, **OLAP**) trata de las herramientas y las técnicas para el análisis de los datos que pueden dar respuestas casi instantáneas a las consultas que soliciten datos resumidos, aunque la base de datos sea extremadamente grande.

El campo del **análisis estadístico** también se incluye en la ayuda a la toma de decisiones. Los lenguajes de consulta de BD no resultan adecuados para el rendimiento de los análisis estadísticos detallados de los datos. Se han creado una serie de paquetes que ayudan en el análisis estadístico. A estos paquetes se les ha añadido interfaces con las bases de datos para permitir que se almacenen en la base de datos grandes volúmenes de datos y se recuperen de forma eficiente para su análisis.

Las técnicas de búsqueda de información intentan descubrir de manera automática las reglas y las pautas estadísticas de los datos. El campo de la **minería de datos** combina las técnicas de búsqueda de la información creadas por los investigadores en inteligencia artificial y los expertos en análisis estadísticos con las técnicas de implantación eficiente que permiten utilizarlas en bases de datos muy grandes.

Las grandes empresas tienen varios orígenes de datos que necesitan utilizar para tomar decisiones

empresariales. Para ejecutar de manera eficiente las consultas sobre datos tan diferentes, las empresas han creado **almacenes de datos**. Los almacenes de datos reúnen los datos de varios orígenes bajo un esquema unificado en un solo sitio. Por tanto, ofrecen al usuario una sola interfaz uniforme para los datos.

### 1.6.3. Almacenes de datos (Data Warehouse)

Los almacenes de datos pretenden dar soluciones a los problemas inherentes de las bases de datos operaciones, de los sistemas de información de gestión y de los sistemas de soporte a la decisión. Convierten los datos operacionales en herramientas capaces de dar base al análisis y toma de decisiones, y se desarrollan aplicaciones de acceso a datos con características basadas en los sistemas **DSS y MIS**.

A través de los mecanismos ETL, se obtiene un almacén de datos para su ulterior explotación: análisis OLAP, minería de datos, visualización de datos, informes, control de eventos, etcétera.

Un almacén de datos es un conjunto de **depósitos de información** reunida de varios orígenes, almacenada bajo un esquema unificado en un solo sitio. Su almacenamiento es duradero, convirtiéndose en datos históricos, y proporciona una interfaz para desarrollar consultas de ayuda a la toma de decisiones. Esta nueva arquitectura combina potentes herramientas de modelado multidimensional con herramientas de acceso a bases de datos operacionales en busca de la causa de esos datos. Supone una vista lógica unificada de los datos.

El almacén de datos se considera como un repositorio lógico central independientemente de su organización física, con diferentes niveles que contienen resúmenes de datos de niveles inferiores. Estos **resúmenes** sólo contienen los datos relevantes para las tomas de decisiones.

Los procesos de extracción, de transformación y de carga son los mecanismos en que los datos se registran en los almacenes de datos y se conocen con las siglas **ETL** (Extract, Transform y Load). Estas operaciones son relativas a las técnicas para la obtención de la información, los mecanismos de transformación para que puedan almacenarse y los procesos de carga en el almacén de datos final. Las herramientas de entrada a estos procesos son los ficheros planos, Flat Files, los registros devueltos por consultas SQL, la gestión de relaciones con los clientes o **CRM** y las bases de datos que disponen los programas que permiten organizar las distintas áreas operacionales de una organización o **ERP**.

Los componentes de un almacén de datos componen una estructura bastante compleja, pero se puede resumir en los siguientes:

**Modelo de datos corporativos:** los esquemas de cada base de datos en particular deben integrarse obteniendo un único modelo de datos que identifique y estructure los requisitos de información del almacén de datos. Se obtiene un único modelo conceptual.

**Limpieza y carga de datos operativos:** se reconoce los sistemas OLTP que intervienen, se corrigen errores de inconsistencia, redundancias inútiles, se definen tipos de datos y longitudes, el dominio de los atributos, se documenta y se cargan desde las bases de datos hacia el almacén.

**Almacén de datos:** contiene todos los datos actuales en un momento determinado. Es necesario equilibrar el alto coste de cargas operativas con la actualización de los datos.

**Extracción y recuperación de datos:** mecanismos de acceso a datos, su visualización, análisis y resultados en

informes. Se usará para ello la base de datos y el diccionario de datos. El análisis multidimensional permitirá realizar consultas complejas.

**Metadatos:** que describen la estructura de los datos y algoritmos para crear los resúmenes.

Independientemente de estos componentes es necesario establecer la arquitectura física del almacén de datos. Esta puede ser de diferentes tipos:

**Arquitectura centralizada:** se usa un único servidor para guardar todo el almacén de datos. Maximiza la potencia de cálculo y facilita el mantenimiento, pero las consultas pueden consumir muchos recursos. La inseguridad aumenta.

**Arquitectura distribuida:** los datos del almacén se reparten entre varios servidores. Se suelen distribuir por temas lógicos. La carga está distribuida, lo que mejora la ejecución de las consultas, pero el mantenimiento se hace más complejo.

**Distribución por niveles:** se eligen diferentes servidores en función del nivel de detalle de los datos que contienen, datos de detalle, datos resumidos y datos muy resumidos. Los datos muy resumidos se pueden repartir en varios servidores para aumentar la disponibilidad de los datos. El acceso a los datos que más se requieren es muy rápido sin penalizar las consultas a los datos de detalles.

La arquitectura más usada es la de tres niveles, que produce un flujo de datos bien organizado. En el nivel más inferior está el servidor de bases de datos, unido a las numerosas fuentes para uso **Front-End**. En el nivel medio se incluye un servidor para el procesamiento analítico en línea, OLAP, y compone el almacén de datos centrado en OLAP. En un tercer nivel se incluyen las herramientas y la interfaz de programación que se usan para el análisis de datos, consultas e informes.

#### 1.6.4. Minería de datos

La minería de datos, o **Data Mining**, es la técnica que recopila datos en la búsqueda de reglas y estructuras a partir de los datos localizados en un almacén de datos. El objeto es la búsqueda del conocimiento en las bases de datos. La plataforma con la que trabaja la minería de datos es el almacén de datos, y ofrece una serie de técnicas que automatizan, en la medida de lo posible, la extracción de información oculta en ella, trabajando bajo las siguientes líneas de actuación:

- Se pretende construir patrones o categorías.
- Se incluye el análisis de información aplicado en los sistemas de soporte a la decisión.
- Se automatizan los procesos de planteamiento y descubrimiento de hechos e hipótesis.
- Se pretende predecir futuras tendencias y comportamientos.
- Se buscan relaciones y patrones globales ocultos en los almacenes de datos.

El procedimiento para resolver un problema a través de la minería de datos se divide en dos grandes etapas: la preparación de los datos y la minería de datos propiamente dicha.

**Preparación de los datos:** se parte del planteamiento del problema definiéndolo y determinando los recursos,

las fuentes de información y su disponibilidad. Se seleccionan las fuentes válidas, realizando un tratamiento y estructuración de la información, eliminando la información no útil y detectando otra que falta. Se unifican criterios de presentación, se eliminan redundancias y duplicados..

**Minería de datos:** se seleccionan las técnicas, los algoritmos en búsqueda de patrones, se seleccionan los modelos y parámetros apropiados. Se interpretan los resultados en busca de modelos útiles en la toma de decisión y se evalúan usando diferentes casos de aplicación.

La minería de datos parte de almacenes de datos u otros almacenes más específicos llamados Data Marts. Las técnicas más usadas son:

**Análisis multidimensional:** se cruzan datos de múltiples formas y con distintos niveles de agregación. Se basan en las bases de datos multidimensionales y técnicas OLAP

**Agentes inteligentes:** análisis para detectar patrones y relaciones. Se usan técnicas con metodologías basadas en sistemas expertos y el aprendizaje automático. Los algoritmos usados son muy avanzados, que usan estructuras complejas como pueden ser las redes neuronales y los árboles de decisión, y avanzadas técnicas tales como el agrupamiento y la lógica difusa. Estas técnicas se usan en gran medida para la detección de patrones.

**Consultas e informes:** son las metodologías tradicionales a través de consultas, y emplean la estadística para el análisis multidimensional y agentes inteligentes. Son metodologías muy útiles para bases de datos relacionales.

**Detección de alarmas:** se ejecutan agentes para desencadenar acciones extraordinarias.

### Bases de datos multidimensionales y análisis multidimensional

En las bases de datos multidimensionales no se usan estructuras bidimensionales, sino que se emplean estructuras más complejas n-dimensionales. Los **hipercubos** son las estructuras más usadas. Los datos tienen una estructura matricial de más de dos dimensiones, permitiendo realizar consultas más complejas. El análisis que se aplica, análisis multidimensional, recae sobre diversos componentes, sean cuantitativos o cualitativos. Cada componente que estudiar supone una dimensión y sus valores se denominan **atributos**. Además, cada dimensión se puede descomponer en otros niveles de detalles.

Para llevar a cabo el análisis multidimensional es necesario disponer de todo el conjunto de datos, es decir, del esquema multidimensional, soportado por bases de datos multidimensionales. Existen mecanismos que transforman el esquema multidimensional en esquemas soportables por sistemas gestores de bases de datos relacionales, por ejemplo, OLAP para bases de datos relacionales. Los esquemas de transformación más conocidos son el **esquema estrella** y el llamado **copo de nieve**.

### Modelo de datos multidimensional

Es un modelo alternativo al modelo entidad-relación, supone un modelado tanto estático como dinámico y está basado en **estructuras multidimensionales**. El fin último de este tipo de modelos es conseguir mayor rendimiento, tiempo de respuesta óptimo y mayor tolerancia a los cambios.

Este modelo tiene los siguientes componentes:

- Atributos de hecho o de medida: atributos supeditados a una función estadística.
- Funciones resumen: las funciones antes mencionadas.
- Atributos de dimensión: los atributos que son cualitativos.
- Series temporales: secuencia de datos sujeta al tiempo.
- Esquema de hecho o de cubo: objeto para analizar.
- Dimensiones: cada eje en el espacio multidimensional.
- Jerarquías: hace referencia a los atributos de dimensión que ofrecen jerarquía a la estructura.

### **Procesamiento analítico en línea (OLAP)**

**OLAP** (On-Line Analytical Processing) es una técnica para bases de datos, generalmente multidimensionales y de gran tamaño, que usa metodologías de síntesis de los datos a través de resúmenes que se ofrecen a niveles superiores.

Las bases de datos que usa OLAP pueden ser multidimensionales o relacionales. Atendiendo a este uso se obtienen los siguientes tipos de procesamiento:

- MOLAP, Multidimensional-OLAP: los datos se almacenan en un sistema de hipercubo en donde cada eje es una dimensión y se usan procesos de modelización para transformar los esquemas relacionales a multidimensionales.
- ROLAP, Relational-OLAP: los datos se almacenan en tablas y se trabaja sobre bases de datos corporativas. Hay mayor coste espacio temporal, ya que se debe transformar el esquema relacional al manejo multidimensional y jerárquico. Su uso está encaminado en el desarrollo de técnicas de almacenamiento y optimización de consultas.
- HOLAP, Hybrid-OLAP: se almacenan resúmenes y datos básicos en memoria, y otros resúmenes en las bases de datos relacionales. Se implementa como cliente- servidor, donde el servidor contiene la base de datos relacional y los cubos HOLAP, y los clientes obtienen vistas de los datos que ofrece el servidor.

### **1.6.5. Big Data**

La recolección masiva de datos a la que antes se hizo mención en la minería de datos. Forma parte de lo que se conoce como **Big Data**. El tamaño de estos almacenes de datos es tan grande que los procesos de análisis en busca de patrones ocultos requieren de técnicas muy avanzadas en su análisis. El objetivo es más desafiante: la obtención de conocimiento en tiempo real, permitido por la propia expansión de internet, avances en las tecnologías de procesamiento de información y el almacenamiento en la nube. Esta unión de almacenes de datos gigantes y algoritmos de análisis en tiempo real es el Big Data.

Los Big Data tienen las siguientes características, las cinco uves:

- **Valor:** aportar valor a la sociedad ayudando a impulsar el desarrollo, la innovación y la competitividad, pero sin olvidar la calidad de vida de las personas..
- **Variedad:** conjugar y combinar cada tipo de información y su tratamiento específico para alcanzar un todo homogéneo.
- **Velocidad:** el análisis de la información disponible debe ser en un tiempo lo más real posible, de lo contrario, su evolución estaría siempre por encima de su tratamiento, haciendo de su explotación tecnologías improductivas, pues se alejan de estados reales.
- **Veracidad:** los datos deben ser de calidad y estar siempre disponibles.
- **Volumen:** la información ha de ser lo más completa posible para que la toma de decisión sea óptima.

Los datos almacenados en los Big Data son de tres grandes categorías: los datos estructurados, los no estructurados y un nivel intermedio.

**Datos estructurados:** la naturaleza de los datos permite estructurar los datos en elementos de almacenamiento como listas, tablas, objetos, etcétera.

**Datos no estructurados:** no es posible localizar un formato que unifique los datos almacenados por los que se almacenan tal como entran en el sistema. Pueden ser planos, por ejemplo, los textos en redes sociales, o de tipo no texto como una fotografía, un video o un audio.

**Datos semiestructurados:** no están almacenados en estructuras de datos, aunque sí se pueden organizar para su tratamiento. Ejemplos de estos son documentos HTML, XML, JSON o bases de datos NoSQL

Un ejemplo de producto comercial para la gestión completa de almacenes de datos gigantes es Apache Hadoop, que se ha convertido en uno de los mejores entornos de trabajo para programar aplicaciones distribuidas de acceso a tecnologías Big Data. Usa un sistema de archivo distribuido denominado HDFS (Hadoop Distributed File System), y está compuesto por nodos de datos y un clúster de datos HDFS.

#### **1.6.6. Bases de datos en la nube**

Los modelos de implantación de bases de datos en la nube son principalmente dos:

**Bases de datos en la nube:** la gestión es análoga a la que se haría si la base de datos estuviera alojada físicamente en la intranet de la organización, pero el sistema está alojado en máquinas virtuales y soportes de almacenamiento que se alquila a un proveedor de servicios en la nube. El modelo que usan los desarrolladores se llama DevOps. La organización supervisa y gestiona la base de datos.

**Base de datos como servicio (DBaaS):** la base de datos se ejecuta en la nube y el proveedor se encarga de las tareas operativas, de mantenimiento, de gestión y de administración, y todo en tiempo real, con la mayor seguridad y alta disponibilidad y con procesos de automatización de software.

Esta nueva tecnología incluye una serie de ventajas que la hacen muy potente:



**Comercialización más rápida:** debido a que el despliegue tecnológico local no es necesario, el producto en la nube entra en desarrollo en tiempos mucho más cortos.

**Costes más bajos:** debido al contrato por uso y la escalabilidad dinámica, el cliente paga según los momentos de alta o baja funcionalidad de su producto en la nube según la demanda. Incluso los servicios pueden ser desactivados en los momentos necesarios. Esto conlleva ahorros muy elevados.

**Despliegue más rápido:** la configuración en la nube es mucho más rápida, simplificando las pruebas y las nuevas implantaciones.

**Riesgos más bajos:** los errores humanos y los esquemas de alta seguridad son responsabilidad del proveedor del servicio, que automatiza, en la medida de lo posible, estos procesos.

Las **tecnologías en la nube** permiten que existan bases de datos autónomas, de tal modo que las organizaciones no tienen que preocuparse de la contratación de administradores de bases de datos. Junto a estas aparece una serie de opciones en la gestión de las bases de datos en la nube:

**Autogestionadas:** la base de datos se ejecuta en la nube, pero la organización la gestiona usando sus recursos internos sin automatización.

**Automatizadas:** la organización colabora con operaciones de ciclo de vida a través de API de servicios en la nube, pero mantiene los accesos y configuraciones de los Servidores de bases de datos y sistemas operativos.

**Autónomas:** se usa automatización y aprendizaje automático. El sistema se encarga de gestionar y ajustar el rendimiento de la base de datos.

**Gestionadas:** es análoga a la automatizada, pero los clientes no acceden a los servidores de bases de datos ni instalan su propio software. La configuración recae en el proveedor de servicio.

El nuevo paradigma sobre la explotación de bases de datos en la nube conlleva nuevos esquemas en el uso de las tecnologías actuales y emergentes. A continuación, se indican algunas de ellas:

- Las cargas de trabajo OLTP se basan en modelos de datos diferentes de los utilizados en las cargas de trabajo OLAP.
- Los datos del tipo documento o multimedia se basan en formatos como XML y Javascript Object Notation (JSON).
- La tendencia actual es hacia la base de datos multimodelo.
- Los modelos DBaaS son más robustos, lo que afianza la transición a bases de datos autónomas en la nube.
- Las plataformas de bases de datos como servicio (dbaaS) es una oferta comercial de administración de bases de datos o almacén de datos diseñado como un servicio de suscripción multiusuario, escalable y automatizado, respaldado por un proveedor de servicios en la nube (CSP).

#### 1.6.7. Ejemplos de tecnologías Business Intelligence

La inteligencia de negocios o **Business Intelligence (BI)** es un conjunto de arquitecturas técnicas, aplicaciones, datos, estrategias, productos y tecnologías enfocadas a la creación y administración de conocimiento, a través del análisis de los datos almacenados en grandes almacenes de datos y que facilita la toma de decisiones. Por ejemplo, los **table-ros** o **Dashboards** son interfaces gráficas que permiten leer los datos recopilados de una forma cómoda y flexible a través de gráficas y tablas. Muestran resúmenes, tendencias clave, comparaciones, excepciones en históricos, entre muchos otros, con el fin de tomar decisiones óptimas. Otra herramienta, los informes ad hoc, dan respuestas a preguntas comerciales de interés. Se basan en los esquemas DSS estudiados con anterioridad y se enmarcan como productos comerciales. También se usa comercialmente la herramienta denominada **descubrimiento de datos**. Esta consiste en la búsqueda de patrones en bases de datos gigantes.

Las tendencias en Business Intelligence actuales son:

- **Tecnología de Analytics:** software de minería de datos e inteligencia artificial (Data Mining, Deep Learning y Machine Learning). Ejemplos: Apache Spark, Azure, IBM Watson Analytics, Oracle Analytics.
- **Tecnología de Big Data:** software para manejar grandes volúmenes de datos provenientes de diversas fuentes y que se generan a gran velocidad siendo necesario su tratamiento en tiempo real. Ejemplos: Apache Hadoop, AdoreBoard, CrownEdition, Domo, Panoply, Text Emotions, Teradata, Tibco.
- **Tecnología del lenguaje:** software de procesamiento del lenguaje natural (Natural Processing Language, NPL). Ejemplos: Apache Lucene and Soir, Apache OpenNLP, Apache UIMA, Azure, Affectiva, DeepFace, GATE, Natural Language Toolkit, Kairos, Sight Corp, Sky Biometry.
- **Tecnologías Integradoras:** integran más de un tipo de tecnología. Ejemplos: SAP BI, Power BI.

A continuación, se indican algunos ejemplos de softwares sobre BI:

- **TAPADID:** es el acrónimo de las bases de las líneas en el análisis de negocio: tecnología, análisis como servicio, predicción de modelado, automatización, minería de datos, percepciones y toma de decisión.
- **Analytics as a service - análisis como servicio:** es un servicio de rastreo que analiza redes sociales con el fin de obtener inteligencia de negocio a partir de los datos obtenidos en la monitorización. Los datos no son tratados por la organización, sino usando un servicio: el análisis como servicio.
- **ERP SAP:** compuesto por excelentes módulos pioneros en el mercado ERP. Los módulos más importantes que ofertan son:
  - **Módulo SAP QM, Quality Management.** Gestiona la información asociada a la calidad.
  - **Módulo SAP MM, Materials Management.** Se usa en el área logística para reducir costos y aumentar ganancias con estrategias de compras. Gestiona la adquisición de bienes y contratación de servicios: stock, facturas, entrada de mercancías. Procesa una solicitud de pedido, la recepciona, la convierte en pedidos a proveedores y registra su factura; en reglas generales, manejo de solicitud de pedidos, gestión y comparación de cotizaciones, realización de pedidos de compra y tramitación de facturas.
  - **Módulo SAP PP.** Planificación de la producción. Gestión de la producción.
  - **Módulo SAP FI.** Para la gestión completa de las finanzas.
  - **Módulo SAP SD, Venta y Distribución.** Gestiona todos los procesos asociados a las ventas y

distribuciones de bienes y servicios: manejo de pedidos, realización de cotizaciones, procesos de órdenes de ventas y trámite de facturas.