

# UD1 – Business Intelligence

## Índice

1.	Introducción al Business Intelligence .....	1
1.1.	Analizando Business Intelligence .....	2
1.2.	Historia del Business Intelligence .....	3
1.3.	Cambios en el procesamiento debido al Big Data .....	5
2.	Modelos de negocio de Business Intelligence .....	7
2.1.	Hechos y dimensiones .....	7
2.2.	Tipos de modelo de negocio .....	9
3.	El modelo KDD (Knowledge Discovery in Databases) .....	12
3.1.	KDD vs Data mining .....	12
3.2.	Etapas del KDD .....	13
4.	Implantación de modelos BI .....	15
4.1.	Técnicas de validación de modelos BI .....	17
4.2.	Test que se realizan en BI .....	17
5.	Arquitectura típica del Business Intelligence .....	19
5.1.	Arquitectura típica del Business Intelligence .....	19
5.2.	Cuadros de mando .....	20
5.3.	Arquitectura BI: estructura tiempo real .....	21
5.4.	Ejemplo BI industrial .....	23
6.	Bibliografía .....	25

# 1. Introducción al Business Intelligence



“El volumen de datos que gestionan las empresas ha pasado de **1,45 Petabytes** en 2016 a **14,6 Petabytes en 2021**, es decir, una cantidad 10 veces mayor en solo un lustro.” Según la encuesta anual de Dell realizada a más de 1000 responsables de toma de decisiones en todo el mundo.

Fuente: <https://www.delltechnologies.com/es-es/data-protection/gdpi/index.htm>

Big Data está cambiando la forma en que las empresas hacen negocios. Los datos están creciendo a un ritmo explosivo, entrando en la empresa de diferentes áreas y en innumerables formatos. Medios de comunicación social, datos de sensores, coordenadas espaciales y fuentes de datos externas como PLC u otros PC, son solo algunas de los nuevos vectores de datos que cualquier empresa debe abordar. El resultado es que la analítica y Las prácticas de Business Intelligence (BI) deben repensarse en el contexto de Big Data.

Muy ligado a los conceptos que hemos visto hasta ahora, encontramos las herramientas Business Intelligence, estas herramientas nos permitirán extraer información de un origen de datos para ser tratadas y presentadas.

Por tanto, el uso de los datos, su transformación en datos útiles para tomar decisiones es lo que se conoce como **Business Intelligence (BI)**. Con esto, podemos obtener una definición más concreta de lo que es BI:

*Business Intelligence es el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar las mejores decisiones en base a datos analizados.*

Las aplicaciones de **Business Intelligence (BI)** son herramientas de soporte de decisiones que permiten en tiempo real, acceso interactivo, análisis y manipulación de información crítica y así ser capaces de generar conocimiento mediante almacenamiento y análisis de información, al servicio de una empresa.

El propósito del Business Intelligence es poder **generar conocimiento de valor que ayude a impulsar al negocio**, gracias a los datos recopilados por diferentes medios. Dicho conocimiento muestra el comportamiento presente y realiza previsiones de futuro, permitiendo corregir errores y tomar decisiones certeras basadas en tal información.

Los usuarios son capaces de acceder y analizar una vasta cantidad de información y analizar sus relaciones y entender las tendencias que últimamente están apoyando las decisiones de los negocios.

Estas herramientas previenen una potencial pérdida de conocimiento dentro de la empresa que resulta de una acumulación masiva de información que no es fácil de leer o de usar

Para tal fin, se requiere de un software Business Intelligence o de herramientas de inteligencia de negocios capaces de extraer datos desde diferentes fuentes y transformarlos en información útil. Por ejemplo, indicadores económicos y de producción que muestren resultados precisos a disposición de los directivos de las empresas.

### 1.1. Analizando Business Intelligence

El éxito de la gestión empresarial pasa por los responsables que toman decisiones tengan toda la información real y verificada que el proceso de negocio pueda proporcionar, con esa información es posible realizar una correcta planificación del propio proceso de negocio.

Cuando se realiza la planificación estratégica se establecen objetivos y esos objetivos representan acciones que deben realizar cada una de las áreas de la organización. **La coordinación entre las diferentes áreas de la organización en cada una de sus actividades garantiza el éxito de los resultados.**

Un área de **producción** no puede funcionar sin tener interacción coordinación de actividades con el área de compras porque es necesario que le suministre los productos de la materia prima etc. El área de comercialización necesita conocer cuáles son los productos que se están fabricando o cuáles son los productos que se están comprando para poder establecer también las estrategias de ventas hacia ese tipo de producto.

El área de **marketing** necesita conocer cuáles son los productos que se están fabricando por lo que se están comprando para poder ofrecer armar las campañas necesarias para poder vender esos productos conocer el stock que va a tener etc.

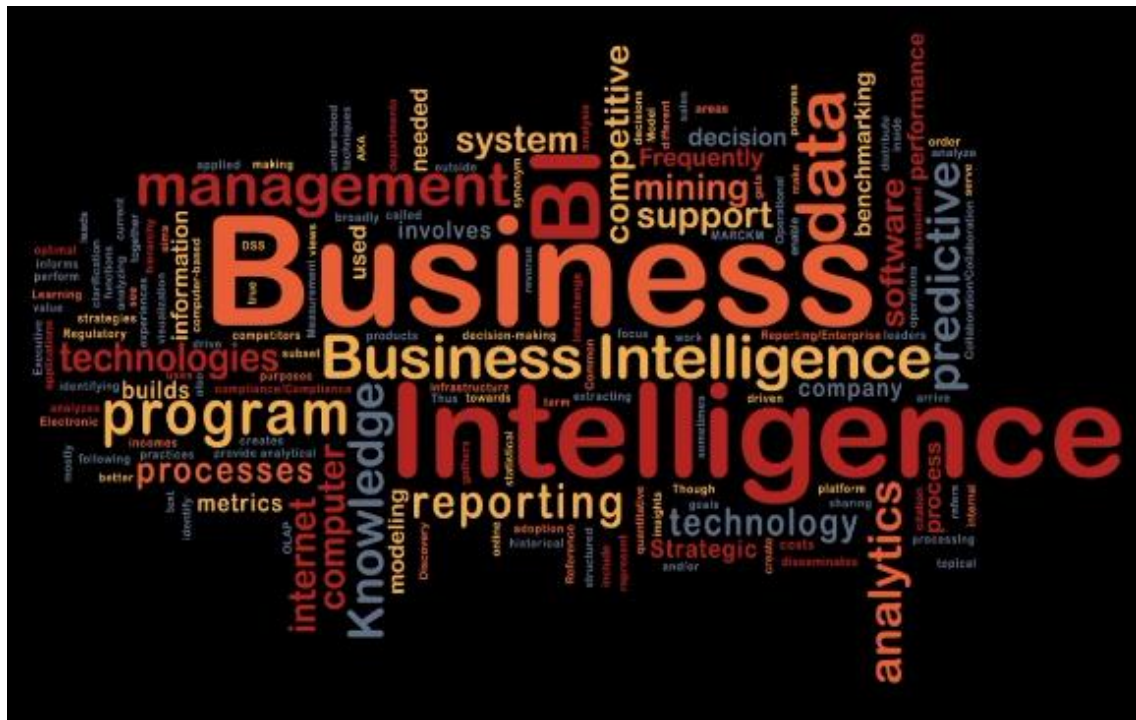
Para que se produzca una coordinación entre las distintas áreas, debemos contar con una herramienta que pueda proporcionar información de manera transversal a todas las partes de la organización, de forma que los responsables puedan tomar la decisión adecuada.

Otro punto importante del negocio es conocer cómo está reaccionando el cliente antes las acciones de marketing y de los productos que se fabriquen. Por otro lado, saber el perfil del cliente, cuáles son los hábitos de consumo, conocer cuál es el grupo target al que va dirigido un determinado producto hace que los responsables puedan potenciar de alguna manera los productos hacia los clientes.



Este tipo de herramientas, que proporciona información empresarial para que los responsables puedan tomar la decisión general se llama Business Intelligence.

Esta disciplina, por llamarla así, se orienta a la recolección e identificación de datos. Dichos datos representan interacciones entre personas, departamentos, productos y clientes, los cuales pueden venir desde dentro de la organización como desde fuera.



Como hemos visto en la UT1, todos estos datos pasan por un proceso de transformación para convertirse en información útil que permite de manera rápida **identificar patrones** que está sucediendo en ese proceso en específico para transformarse en conocimiento sobre cómo se están realizando las actividades.

## 1.2. Historia del Business Intelligence

Históricamente, la tecnología de Business Intelligence ha encontrado lugar en dos niveles primarios:

- Entre los altos ejecutivos quienes necesitan obtener información estratégica
- Entre los administradores de la línea de negocios que son responsables del análisis táctico.

Estas tradicionales actividades de soporte a la decisión son importantes, pero ellos solamente muestran superficialmente el potencial de la inteligencia de negocios dentro de la empresa, involucrando quizá el 5% de los usuarios y el 10% de los datos disponibles.

Al principio, las organizaciones dependían de sus departamentos de sistemas de información para proporcionarles informes estándar y personalizados. Esto ocurrió cuando la mayoría de los usuarios no tenía acceso directo a los ordenadores.

Esto cambió en los años 70's cuando los sistemas basados en servidores aparecieron. Aun así, estos sistemas eran usados principalmente para transacciones de negocios y sus capacidades de realizar informes se limitaba a un número predefinido de ellos.

Con el paso del tiempo, fueron desarrollados los sistemas de información ejecutiva (EIS, por sus siglas en inglés), los cuales fueron adaptados para apoyar a las necesidades de ejecutivos y administradores.

Con la entrada de la PC, y de computadoras en red, las herramientas de BI proveyeron a los usuarios de la tecnología para crear sus propias rutinas básicas y reportes personalizados.

### **Ventajas de contar con BI**

- **Aumento de la eficiencia:** toda organización recopila información de sus departamentos, con herramientas de este tipo, la recopilación y posterior análisis es más rápido y eficiente.
- **Respuestas inmediatas:** al tratarse de un sistema centralizado, se obtienen tiempos de respuesta muy bajos, a esto hay que añadir que, estos sistemas están preparados para trabajar con grandes volúmenes de datos.
  - Ejemplo: ventas en plataformas online y relación entre mostrador y almacenes repartidos por un territorio
- **Información precisa:** los datos almacenados son, a todas luces, información real, no subjetiva ni basada en intuiciones, con dicha información, se obtiene una valoración objetiva del punto deseado.
  - Ejemplo: información relativa al posicionamiento de móvil de un usuario (en caso de pérdida)
- **Conocimiento del flujo de negocio:** se analizan las partes del negocio y se obtienen conclusiones respecto a partes específicas de la empresa.
  - Ejemplo: explicación de la caída puntual de producción eólica en una determinada fecha
- **Mejora del control sobre las áreas funcionales de la organización:** cada parte de la organización está relacionada con otras, por lo que, con este tipo de herramientas se puede centralizar y mejorar las áreas importantes o que lo requieran.
  - Ejemplo: reubicación de un trabajador con cierta experiencia a un departamento donde ha habido una bajada de rendimiento.

### **Tendencias y Desarrollos**

Estas son algunas de las tendencias y desarrollos más importantes de la inteligencia de negocios en la actualidad:

- **La combinación de software y servicios de consultoría** – Los proveedores están empezando a ofrecer «información como servicio» y Business Intelligence a los clientes, en lugar de vender el software y la infraestructura que las empresas necesitan para acceder al BI por su cuenta.
- **Aumentar el autoservicio** – El software se centra cada vez más en aumentar el número de funciones que se pueden realizar sin tener que involucrar al personal de TI o a los especialistas en datos.
- **Inteligencia de negocios basada en la nube** – Mientras que la computación en nube se ha afianzado en otras áreas, también está comenzando a ganar impulso en el área de la inteligencia empresarial. A medida que evolucione, permitirá a las empresas utilizar el Business Intelligence sin tener que dedicar recursos internos a la gestión de la infraestructura y a la actualización del software.
- **Inteligencia móvil** – El móvil se convierte en un elemento clave de la actividad diaria y no es diferente en el campo de la inteligencia de negocios. Las herramientas móviles

permiten a los responsables de la toma de decisiones acceder a la información dondequiera que la necesiten, no sólo cuando están en la oficina.

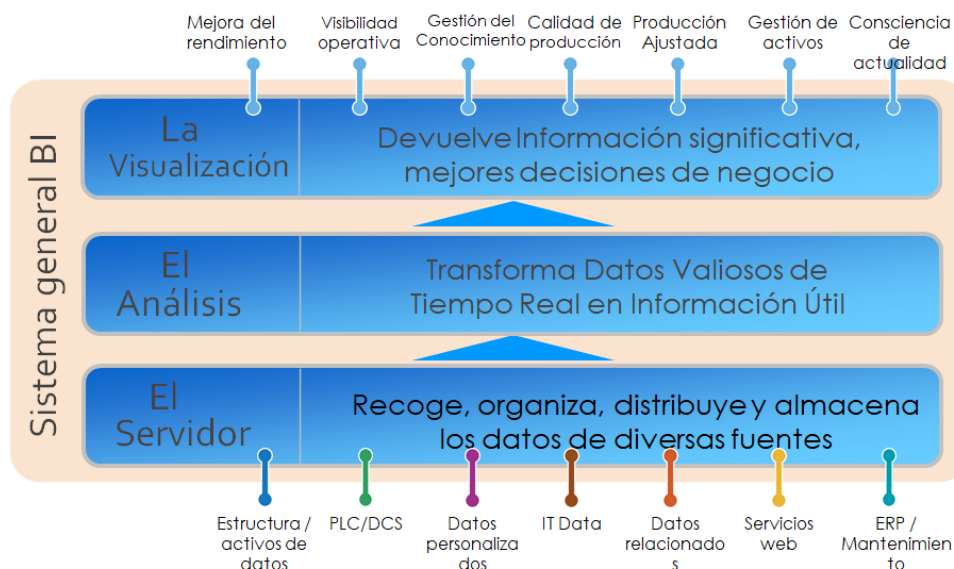
- **Big Data** – Las empresas tienen acceso a más datos que nunca, muchos de los cuales provienen de fuera de la organización en una forma no estructurada. La inteligencia de negocios se combina cada vez más con el análisis de grandes volúmenes de datos, de modo que las empresas pueden tomar decisiones utilizando toda la información que tienen a su disposición, independientemente de la forma que adopte.

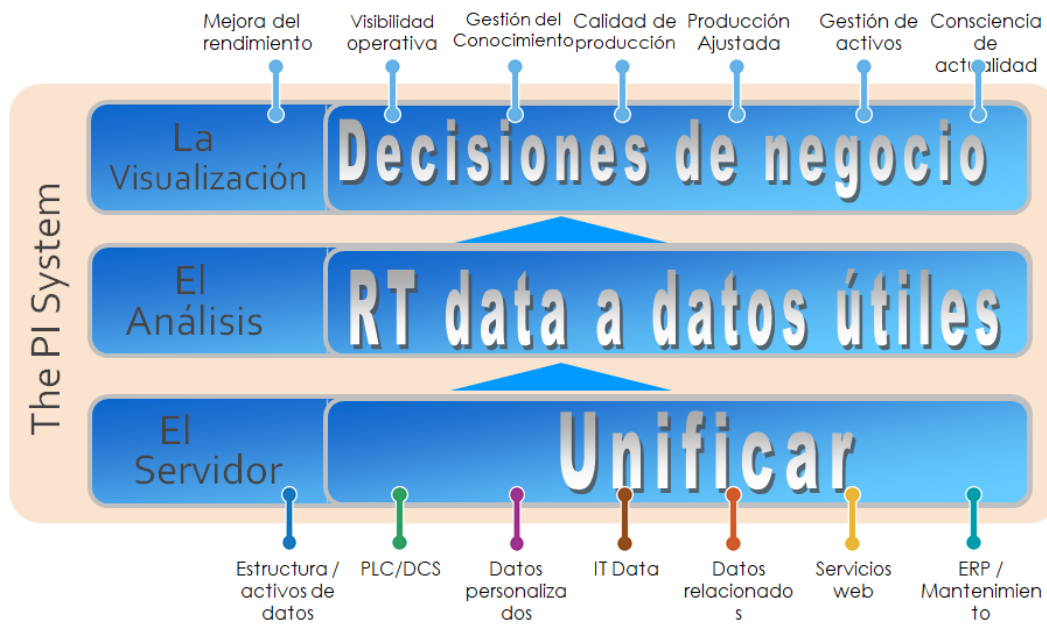
### 1.3. Cambios en el procesamiento debido al Big Data

Tradicionalmente, los grandes conjuntos de datos residirían en un almacén de datos, a menudo gestionados sistemas avanzados de gestión de bases de datos relacionales (RDBMS). No obstante, este tipo de configuraciones no refleja la naturaleza cambiante del Big Data. Hemos de asumir que los datos (y las fuentes desde donde los obtenemos) han cambiado, por tanto, también debe cambiar la forma en que se procesa.

Herramientas de BI tradicionales que se basan exclusivamente en los almacenes, esto implica que dichos sistemas no gestionen de forma eficaz grandes conjuntos de datos que contienen formatos no estructurados y semiestructurados. Para apoyar a Big Los datos, las herramientas modernas de procesamiento analítico deben:

- ✓ Alejarse de las herramientas de BI tradicionales que miran hacia atrás para buscar plataformas analíticas más progresistas.
- ✓ Apoyar un entorno de datos menos centrado en la integridad con solo almacenes de datos corporativos tradicionales y más centrado en una fácil integración con fuentes de datos externas.
- ✓ Apoyar una combinación de datos estructurados, semiestructurados y datos no estructurados sin TI compleja y que consume mucho tiempo esfuerzos de ingeniería/procesamiento.
- ✓ Procesar datos de manera rápida y eficiente para devolver respuestas antes de que se pierda la oportunidad de negocio.
- ✓ Presente al usuario empresarial una interfaz que no requieren amplios conocimientos de TI para operar.





### Ejemplos Big Data

- **Netflix:** utiliza BI para conocer los detalles de los gustos y las preferencias de cada persona según su comportamiento en la plataforma.
  - Predice los comportamientos de conducta de los usuarios, utilizando los datos para recomendar diferentes series y programas en base a los me gusta, el abandono, el dispositivo de visualización, el idioma, los saltos de tiempo, así como los contenidos compartidos a través de las redes sociales.
- **Telco:** cadena de alimentación de Reino Unido, empezó a utilizar el Business Intelligence en su tarjeta de fidelización Clubcard. Analizando el comportamiento de los usuarios, empezó a distribuir cupones descuento.
  - Uniendo datos meteorológicos y el uso de Machine Learning para determinar las ventas de productos, consiguió ahorrar 100 millones de euros en productos de caducidad y a reducir el desperdicio gracias a ser capaz de predecir las ventas por producto de cada tienda.
- **UPS:** Orion es el nombre de su sistema de Inteligencia Artificial.
  - El algoritmo On-Road Integrated Optimization and Navigation define en base a 1.000 páginas de código y 200.000 posibilidades de ruta en tiempo real para que los mensajeros que transportan un producto desde punto A hasta un punto B puedan hacerlo de la manera más rápida posible.
  - Esta iniciativa ha ayudado a la compañía a ahorrar en gasolina y reducir las emisiones de carbono



## 2. Modelos de negocio de Business Intelligence

Cuando se habla de modelos de datos en BI (Business Intelligence), es muy común que nos vengan a la mente las **herramientas** que se utilizan para implementarlo, sin embargo, el concepto de modelos de datos en BI va más allá de esto.

Es necesario seguir una **metodología** y **tipos de modelo específicos** para esta clase de soluciones, que permitan sacar el mejor provecho y beneficio a nuestra solución final.

### Las preguntas del negocio

Para comenzar a realizar un modelo de datos en BI, es **importante conocer el propósito de la información que se va a analizar**, por consiguiente, el negocio.

A partir de allí, se utiliza esta información como base de nuestro **modelo dimensional**, que permita responder las preguntas relacionadas al negocio.

Por ejemplo:

- ¿Cuánto dinero se produce en ventas anualmente?
- ¿Cuál sucursal de mi negocio tiene la mayor venta mensual?
- ¿Cuántos ingresos genera el producto que más vende mi negocio?
- ¿Cuál es mi top 5 de productos que generan las mayores ganancias?
- ¿Cuáles son los empleados con mayores ventas por sucursal?
- Respecto al año anterior, ¿Cómo han variado mis ventas por producto y sucursal?

Estas preguntas nos van a ayudar a delimitar el alcance de nuestro modelo y entender mejor los requerimientos del negocio.

Serán pieza clave para enfocar nuestro modelo de datos en BI hacia un [Data Mart](#) (DM - **almacén de datos orientado a un área específica**) o [Data Warehouse](#) (DW - **almacén electrónico donde generalmente una empresa u organización mantiene una gran cantidad de información**), dependiendo de la magnitud y áreas a considerar.

### 2.1. Hechos y dimensiones

Una vez obtenidas las respuestas a las preguntas del negocio, hemos de implementar un modelo de datos en BI, para ellos, debemos transformar los datos obtenidos en dos conceptos: «**Hechos**» y «**Dimensiones**».

#### Hechos

Los **hechos** en los modelos de datos en BI se los puede definir como **uno o varios eventos dentro del negocio que se requieren registrar y que queremos analizar o medir**, por ejemplo, las ventas.

A diferencia de un **modelo de datos relacional normalizado**, en BI se busca exactamente lo opuesto, es decir, necesitas desnormalizar la información. El propósito final será el de **facilitar y optimizar el proceso de consultas al modelo final**, desde el punto técnico y desde la perspectiva del entendimiento del usuario final.

La tabla de «**hechos**» que resulta de este proceso, debe considerar las métricas en la mínima expresión que se requiera para su análisis posterior (granularidad).



Además, debe incluir identificadores de negocio únicos para relacionarla con las «dimensiones» (atributos de entidades relacionadas al hecho) del modelo.

### Dimensiones

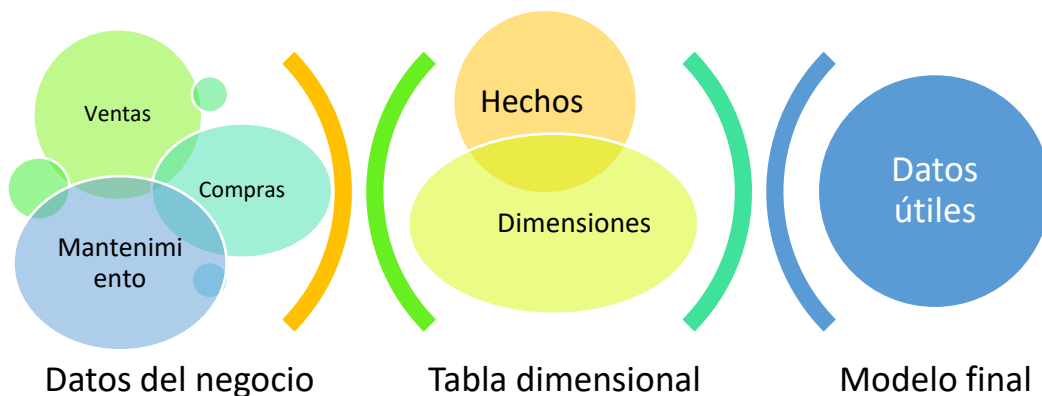
Una dimensión se puede definir como la perspectiva de análisis del hecho del modelo de datos en BI.

Se puede explicar sencillamente, siguiendo el ejemplo de las preguntas del negocio respecto a las ventas.

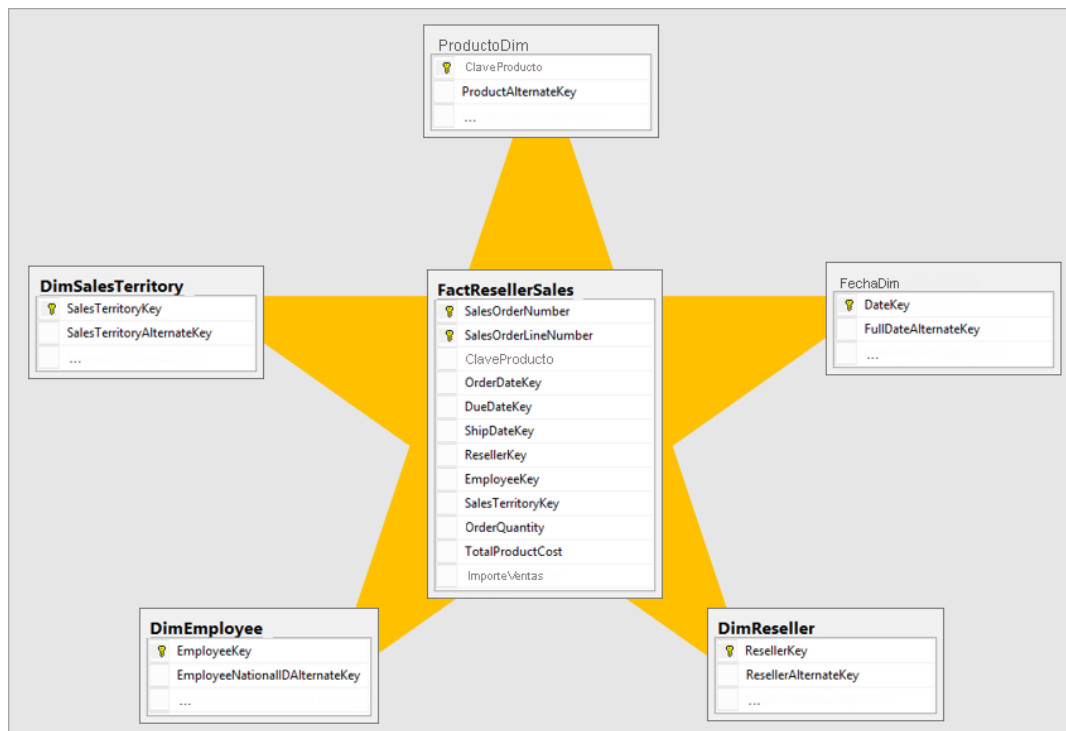
- ¿Cuánto dinero se produce en **ventas** **anualmente**?

La dimensión en esta pregunta tiene relación con el **tiempo** en el que se desarrolla el **hecho**, es decir, es la perspectiva de análisis que se utiliza para evaluar el dinero que se produce en ventas. Dicho de otra forma, la dimensión es un margen temporal en el cual vamos a medir/obtener los hechos.

A diferencia de los hechos, las dimensiones solamente consideran los **atributos de las entidades** que los describen, teniendo **registros únicos y con historia limitada**.



Las columnas de clave de dimensiones determinan la dimensionalidad de una tabla de hechos, mientras que los valores de clave de dimensiones determinan la granularidad de una tabla de hechos. Por ejemplo, imagina una tabla de hechos diseñada para almacenar objetivos de ventas que tiene dos columnas de clave de dimensiones **Date** y **ProductKey**. Resulta fácil comprender que la tabla tiene dos dimensiones. Pero la granularidad no se puede determinar sin tener en cuenta los valores de clave de dimensiones. En este ejemplo, imagine que los valores almacenados en la columna Date son el primer día de cada mes. En este caso, la granularidad está en el nivel mes-producto.

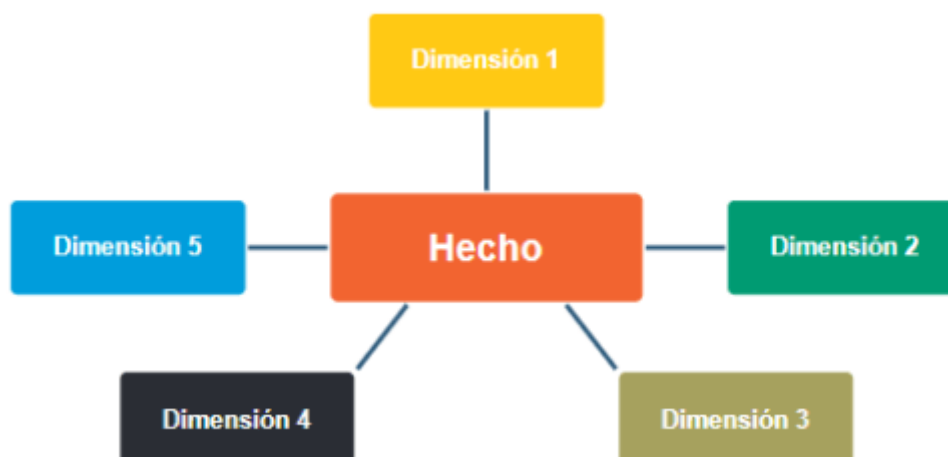


Normalmente, las tablas de dimensiones contienen un número relativamente pequeño de filas. Por el contrario, las tablas de hechos pueden contener un gran número de filas y seguir creciendo con el tiempo.

## 2.2. Tipos de modelo de negocio

Los modelos de datos en BI, se consideran básicamente de dos tipos: Modelo en estrella y Modelo en copo de nieve.

### Modelo de estrella

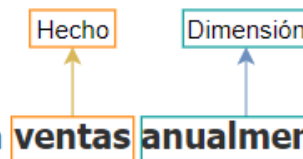


Este tipo de modelo se basa en una tabla central que contiene **el hecho**, concepto de análisis (tabla de hecho), los campos id que la relacionan con las tablas de dimensiones y los cálculos relacionados al hecho en la mínima granularidad requerida.

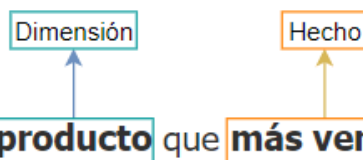
Este modelo se considera el ideal para utilizar en conjunto con herramientas de visualización de datos, ya que facilita el filtrado de los datos. Este tipo de tabla la veremos durante las prácticas en Qlik Sense.

Ahora bien, desglosaremos un poco el concepto con el caso práctico que hemos venido desarrollando. Utilizaremos las preguntas del negocio para entender las dimensiones que necesitamos en relación al hecho de las ventas:

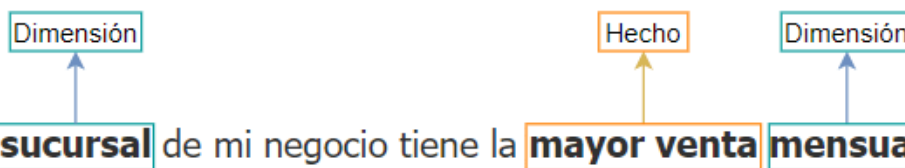
- ¿Cuánto dinero se produce en **ventas** **anualmente**?



- ¿Cuántos ingresos genera el **producto** que **más vende** mi negocio?



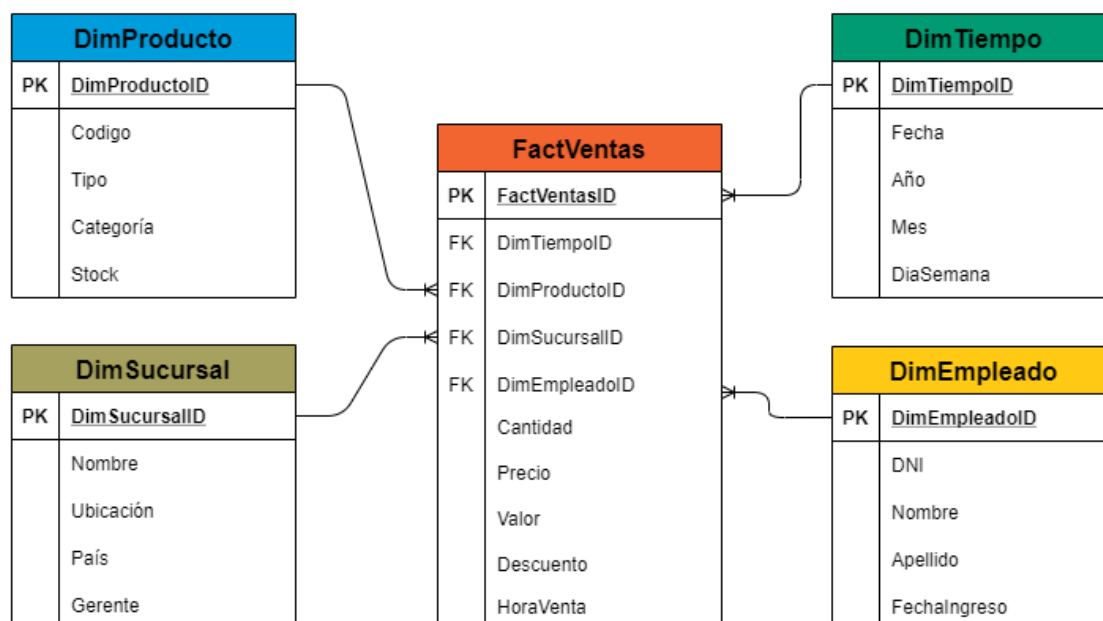
- ¿Cuál **sucursal** de mi negocio tiene la **mayor venta** **mensual**?



- ¿Cuáles son los **empleados** con **mayores ventas** por **sucursal**?



Descomponiendo las preguntas anteriores, podemos obtener la siguiente tabla:



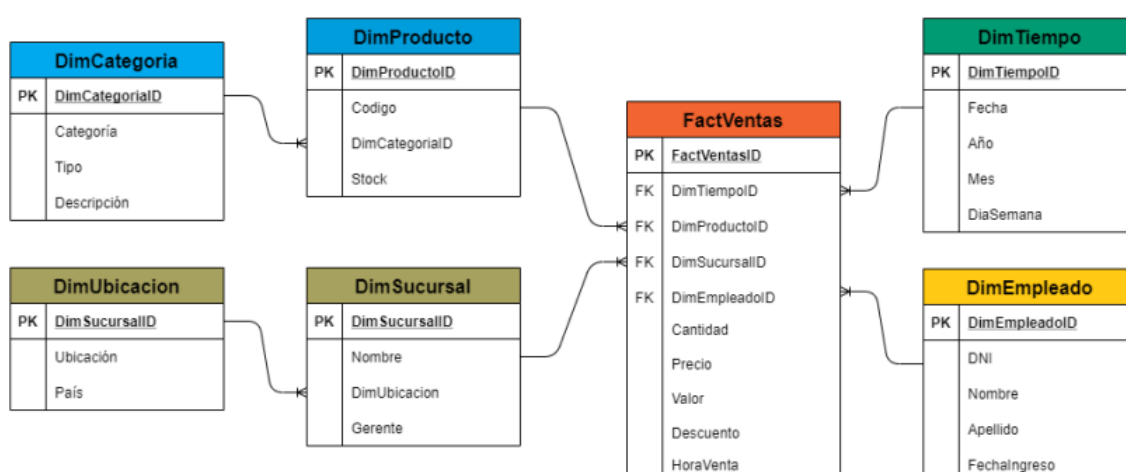
Tal y como podemos apreciar, la tabla central de la estrella posee los atributos principales de las otras, además de elementos propios de su tabla.

### Modelo de copo de nieve

Este tipo de modelo es más complejo que un modelo en estrella, pero el concepto de creación parte de la misma base de análisis.

Se presenta cuando las dimensiones tienen más de una tabla para conformarla, y el objetivo es básicamente disminuir el almacenamiento con la normalización de estas tablas. Bajo este modelo, la tabla de hecho no está relacionada directamente a todas las tablas que componen el modelo de datos.

Siguiendo nuestro ejemplo, podríamos considerar entonces el siguiente esquema de modelado en copo de nieve:



### ¿Cuál modelo de datos en BI debería utilizar?

Cada tipo de modelo tiene sus ventajas y desventajas, y va a depender del escenario de negocio y de disponibilidad de recursos de hardware y software que dispongas, la respuesta a la interrogante.

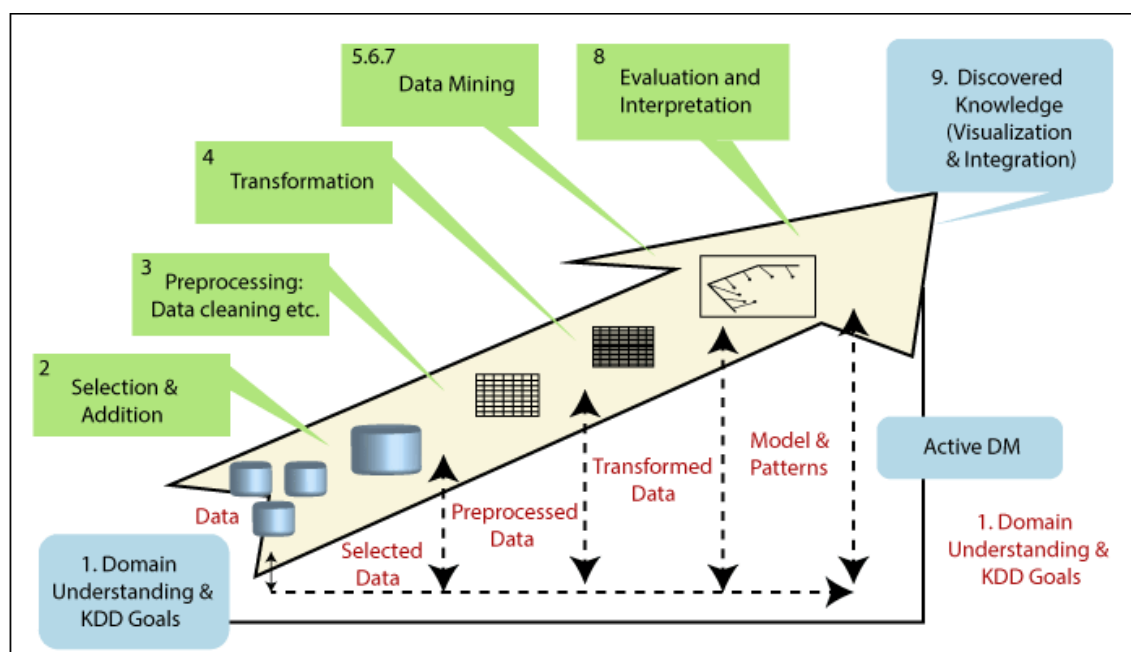
Por ello, te dejamos a continuación una tabla comparativa entre ambos modelos de datos en BI:

	Estrella	Copo de nieve
<b>Mantenimiento</b>	Redundante. Requiere mayor mantenimiento.	No redundante. Requiere menos mantenimiento.
<b>Facilidad de uso</b>	Fácil uso, consultas menos complejas. Requiere poco conocimiento especializado.	Difícil de entender, consultas más complejas. Requiere gran conocimiento especializado.
<b>Joins</b>	Poca cantidad.	Gran cantidad.
<b>Rendimiento de las consultas</b>	Ejecuciones más rápidas.	Ejecuciones más lentas por los cruces de información.
<b>Dimensiones</b>	Una tabla por dimensión.	Más de una tabla en algunas dimensiones.
<b>Nivel de normalización</b>	Desnormalizado.	Tablas de dimensiones normalizadas, tablas de hechos desnormalizadas.

### 3. El modelo KDD (Knowledge Discovery in Databases)

Cuando hablamos de grandes cantidades de datos, el Descubrimiento de Conocimiento en Bases de Datos o KDD se refiere al proceso de identificar patrones válidos, novedosos, potencialmente útiles y principalmente entendibles. Es básicamente un proceso automático en el que se combinan descubrimiento y análisis.

El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice.



#### 3.1. KDD vs Data mining

Uno de los elementos más conocidos del proceso KDD es el de la minería de datos (**Data mining**). Knowledge Discovery implica la evaluación e interpretación de patrones y modelos para tomar decisiones con respecto a lo que constituye conocimiento y lo que no lo es. Por lo tanto, el KDD requiere de un amplio y profundo conocimiento sobre tu área de estudio.

Por otra parte, la Minería de Datos, exploración de datos o Data Mining, no requiere tanto conocimiento sobre el área de estudio, sino más conocimiento técnico.

Como mencionamos anteriormente, la Minería de Datos es un paso que forma parte del KDD e implica el análisis de grandes cantidades de datos observacionales, para encontrar relaciones insospechadas.

### 3.2. Etapas del KDD

El proceso kdd que se muestra en la imagen anterior es interactivo e iterativo, involucra numerosos pasos con la intervención del usuario en la **toma de muchas decisiones**. Se resume en las siguientes etapas:

- Selección.
- Preprocesamiento/limpieza.
- Transformación/reducción.
- Minería de datos (data mining).
- Interpretación/evaluación

#### Etapa de selección

En la etapa de selección, una vez identificado el conocimiento relevante y prioritario y definidas las metas del proceso kdd, desde el punto de vista del usuario final, se crea un conjunto de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento.

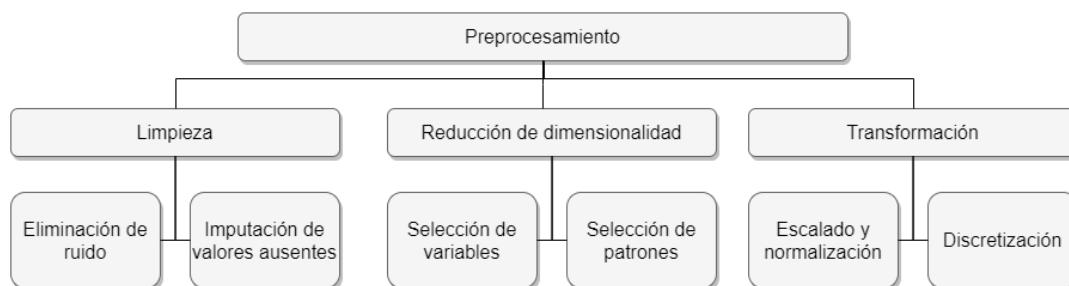
La selección de los datos varía de acuerdo con los objetivos del negocio.

#### Etapa de pre-procesamiento / limpieza

Limpieza de los datos: durante la recogida de datos, ya sea esta manual o automática, así como en su posterior codificación y transmisión hasta llegar a formar parte del conjunto de datos, es habitual que se produzcan errores. Estos se traducen normalmente en dos tipos de problemas. El primero viene dado por datos que son claramente erróneos y, por tanto, introducirían ruido (noisy data) en el proceso de extracción de conocimiento en lugar de resultar útiles. El segundo estriba en la pérdida de algunos datos (data mining), casuística que dejaría huecos en el conjunto de datos. Las tareas de limpieza de datos, por ejemplo mediante algoritmos de eliminación de ruido o de imputación de valores ausentes, abordarían estos obstáculos.

#### Etapa de transformación/reducción

Una vez que los datos están limpios y no contienen redundancias, aspectos de los que se ocupan las operaciones previas, podría pensarse que ya pueden usarse para el aprendizaje de un modelo. No obstante, hay acciones que podrían mejorar los datos de cara a hacer más efectivo ese aprendizaje. Entre ellos están la normalización, el escalado y la discretización. Se trata de operaciones que transforman los datos originales, casi siempre de manera reversible, produciendo una nueva versión más conveniente para el análisis KDD.



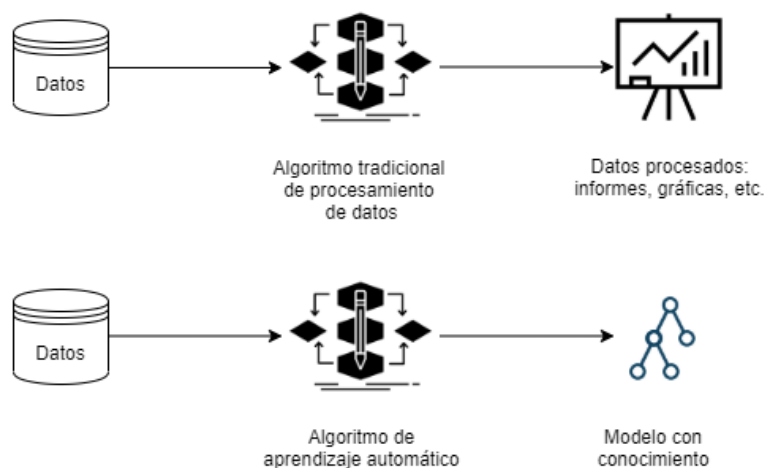
## Etapa de minería de datos

Tras el preprocesamiento, los datos están preparados para la siguiente fase. En esta se usa un algoritmo de minería de datos a fin de extraer de estos el conocimiento que no resulta obvio ni es trivial, aunque esté implícito en ellos.

Existen multitud de métodos que es posible aplicar en este punto, incluyendo diversas técnicas estadísticas, algoritmos matemáticos de optimización y, por supuesto, métodos de Machine Learning.

A diferencia de en los pasos previos, donde se requiere una exploración de los datos que determine qué operaciones es preciso llevar a cabo y la intervención del experto es imprescindible, en la fase de minería de datos es donde suele recurrirse al Machine Learning.

Un algoritmo de aprendizaje automático procesa los datos y lo que genera como salida no son nuevos datos, como es habitual en la mayoría de algoritmos de ordenador, sino un modelo que representa el conocimiento extraído:



## Evaluación y vuelta atrás

Como puede apreciarse en el esquema que resume el proceso de KDD, prácticamente todos los pasos cuentan con una conexión hacia atrás que identifica su naturaleza iterativa. Ese retorno a pasos previos tiene múltiples destinos en la fase final de evaluación e interpretación del conocimiento que se ha obtenido.



Dependiendo del tipo de modelo que se haya generado y de cuál sea su finalidad, hay disponibles diferentes conjuntos de métricas que permiten medir cómo de bueno es ese modelo. La mayor parte de esas medidas están acotadas, es decir, facilitan valores entre un intervalo conocido, por lo que es posible determinar si el conocimiento extraído es útil o no de manera inmediata.

En función del rendimiento obtenido, y del análisis efectuado para determinar qué problemas plantea el modelo, sería preciso realizar ajustes en el algoritmo de Aprendizaje Automático (ML) o bien introducir cambios en los pasos previos de preprocesamiento y preparación de los datos.

## 4. Implantación de modelos BI

Con las ventajas que ofrece un sistema BI a las empresas, cada vez es más frecuente que cualquier empresa quiera acceder a este tipo de sistemas, permitiéndoles trabajar de manera ágil, dotando de eficiencia la gestión de la propia empresa y permitiendo tomar decisiones certeras.

Durante la implantación de un modelo BI, podemos definir las siguientes etapas:

- Definir procesos críticos
- Definir roles
- Definir métricas e indicadores
- Identificar fuentes de información
- Definir arquitectura y tecnología
- Diseñar paneles junto a usuarios
- Implementación con metodologías ágiles

### 1. Definir procesos críticos del negocio

Cada organización tiene una estructura organizacional y operativa que la diferencia y define su potencial competitivo.

Por eso, el primer paso para comenzar a trabajar un modelo de Business Intelligence es revisar cómo se organiza una compañía e identificar cuáles son los procesos fundamentales del negocio.

Estudiar los indicadores de control para medir los resultados alcanzados por la gestión dentro de los objetivos y metas trazadas en el tiempo por la empresa, nos darán las luces para obtener cuáles son estos procesos clave.

Generalmente, hay áreas que resultan más críticas y suelen repetirse. Algunas de ellas son: ventas, supply chain, satisfacción al cliente y marketing.

### 2. Definir los roles clave de cada área

Una vez que definimos los procesos críticos del negocio, hay que definir los roles clave dentro de las áreas.

En los proyectos de Business Intelligence se distinguen dos grandes grupos de usuarios. El primero de ellos, corresponde a quienes producen la información, mediante herramientas desktop para la creación de informes o modelos.

El segundo grupo son los consumidores de la información, quienes consultan los reportes que crea el primer grupo, para la toma de decisiones.

### 3. Definir las métricas e indicadores por área de negocio y roles

El tercer paso es definir las métricas e indicadores por cada área. Lo esencial de este paso es que sea un desarrollo en conjunto con las personas que trabajan en cada departamento. De esta forma, el desarrollador podrá recopilar información de los diferentes departamentos y converger los datos de manera más eficiente.

### 4. Identificar las fuentes de información

El cuarto paso a la hora de implementar un proyecto de Business Intelligence es buscar las fuentes de información que contienen los datos.

Los datos que vamos a pesquisar pueden estar contenidos en diferentes fuentes, como, por ejemplo, si son datos asociados de logística, un sistema de WMS; un ERP, planillas Excel, entre otros.

Los datos son la unidad primaria de información, que por sí sola no aporta valor. La información es el conjunto de datos procesados que tienen un significado y, por lo tanto, son de utilidad.

Por último, está el conocimiento que combina la experiencia, valores de una organización y la información. Aquí surge el BI, optimizando la data que poseen las empresas para su disponibilidad de manera fácil y eficaz.

### 5. Definir la arquitectura y tecnología

Una vez identificadas las fuentes de información, se definen la arquitectura y las herramientas que se utilizarán.

La tendencia en la actualidad es incorporar una arquitectura cloud, aunque no necesariamente toda la estructura de un proyecto de BI tiene un servidor en la nube. Puede ser una mezcla de ambos: on premise y cloud.

Junto a esto se definen las herramientas tecnológicas que se utilizarán para la visualización de la data, pueden ser: Power BI, Tableau, Qlikview, entre otros.

### 6. Definir y diseñar los paneles en conjunto con los usuarios

El sexto paso consiste en diseñar los paneles o dashboards con cada una de las áreas y roles que definimos anteriormente.

Generalmente, se trabajan tres tipos de paneles:

- **Estratégico:** los paneles estratégicos muestran los indicadores clave de una compañía y representan el estado de ésta. Este se actualiza constantemente con datos de la organización y monitorea factores críticos de éxito. Este tipo de dashboards permiten que los altos ejecutivos y tomadores de decisión, determinen acciones sobre el rumbo que va a tomar la empresa.
- **Analítico:** el panel analítico está enfocado en los roles que se dedican a realizar análisis de la información. Estos dashboards entregan una visión completa de los datos y permiten identificar tendencias, compararlas con múltiples variables, crear predicciones y objetivos. En síntesis, este tipo de tablero es útil para implementarse en la estrategia de inteligencia empresarial.

- **Operacional:** estos tableros se enfocan en el seguimiento de procesos operativos y monitorean y analizan las actividades de una empresa en un área de negocios específica.

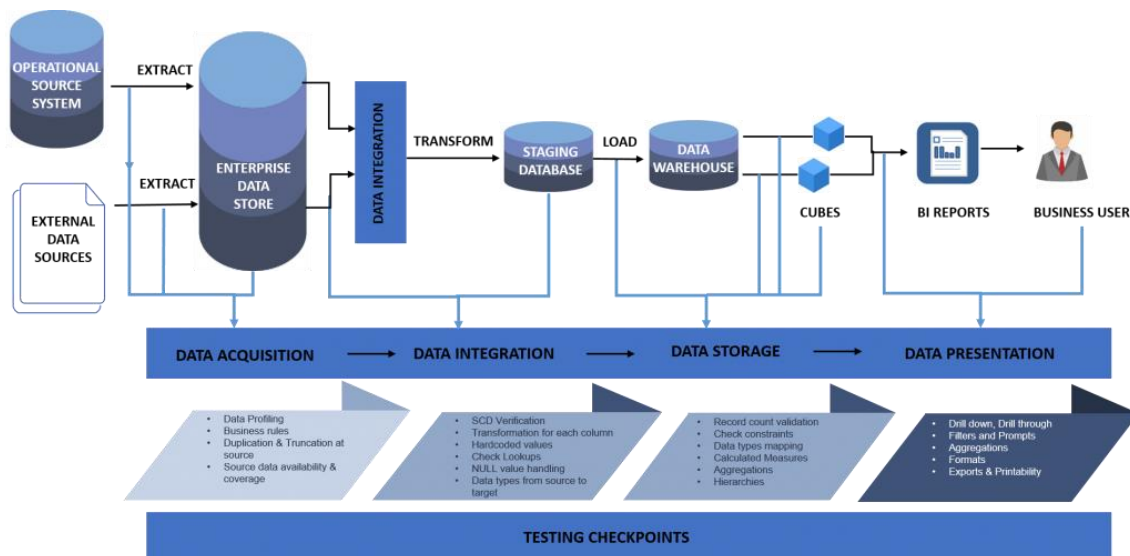
Por ejemplo, en el área de logística, se puede mostrar la productividad de una bodega: cuántos despachos hay en cola, cuál es el tiempo de entrega, etc.

Su función no se limita al control. Por el contrario, sirve para fomentar la proactividad y adelantarse a problemas.

#### 4.1. Técnicas de validación de modelos BI

El testing es una actividad fundamental para cualquier proyecto de software TI. Por ello, se le debe de dar la misma importancia a un proyecto de Business Inteligencie (BI).

Si nos ponemos a repasar todas las etapas que tiene una solución de BI, desde la extracción de datos hasta la disposición final al usuario, podemos ver que hay varias entradas en donde podríamos aplicar técnicas de testing tanto para validación como para control.



Por lo general, los controles que se realizan son en la etapa de extracción, transformación y carga, conocida como ETL (Extract, Transform & Load). Se valida que los procesos no se “caigan”, y enviando avisos si esto sucede. Hay pocos procesos de chequeo de la calidad de los datos. Y pocas veces se realiza testeo rutinario de la estructura. Esto es, verificando que no haya errores que puedan llevar a que la información que se muestra sea incorrecta.

Los tests por lo general en la práctica carecen de una estructura y son más que nada a instinto de lo que debería pasar. Es decir, contrastando los datos que se devuelven a nivel usuario con lo que se encuentra en la base de datos.

#### 4.2. Test que se realizan en BI

Las pruebas generales que aplicamos a cualquier sistema TI en una empresa también son válidas para los sistemas BI:

- **Pruebas unitarias:** son las pruebas que se realizan para validar cada uno de los componentes de una solución. Deben llevarse a cabo durante la etapa de desarrollo. Los

elemento más críticos que deben someterse a este tipo de prueba son la lógica ETL, reglas de negocio y cálculos implementados en la capa de OLAP y la lógica de KPI.

- **Pruebas del sistema:** estas pruebas deben asegurar que se puede construir y desplegar con éxito. Hay que controlar que no surgen problemas durante la ejecución del trabajo. Y por último, confirmar que el sistema actúa del modo esperado con todas las partes constituyentes de la solución funcionando en conjunto.
- **Pruebas de validación de datos:** se trata de validar los datos dentro del Data Warehouse. Esta prueba debería realizarla alguien de negocio, ya que este perfil es quien mejor conoce los datos. Y puede validarlos con mayores garantías de éxito.
- **Pruebas de aceptación de usuario:** se trata de garantizar que los datos que se proporcionen al usuario final cumplen con sus expectativas.
- **Pruebas de rendimiento:** se tratan de validar el rendimiento de la solución en condiciones reales. Para ello, hay que tener en cuenta la arquitectura de datos, la configuración del hardware, la escalabilidad del sistema o la complejidad de las consultas.
- **Pruebas de regresión:** se trata de volver a probar la funcionalidad. Hay que garantizar que el desarrollo implementado no ha tenido impacto negativo en el data warehouse y base de datos. Y que sigue funcionando todo en su conjunto.

De manera más específica, podemos definir las pruebas de un sistema BI basándonos en las anteriores:

### 1 - Verificar los datos en la fuente:

Los datos comerciales generalmente no provienen de una fuente y en un solo formato. Debemos asegurar que la fuente y el tipo de datos que envía coincidan. Ello conlleva realizar una validación de los datos en este primer punto.

Supongamos que los datos de un alumno se envían desde una fuente para su posterior procesamiento y almacenamiento. Asegúrese de que los detalles sean correctos, de forma que sea posible obtener los verdaderos o actualizados en el primer punto.

Esta suele ser la etapa de 'Extracción' del ETL.

### 2 - Verifique la transformación de datos:

En esta fase, los datos llegan sin procesar para su posterior proceso.

- Los tipos de datos de origen y destino deben coincidir. Por ejemplo: no puede almacenar la fecha como texto.
- Las restricciones de clave primaria, clave externa, nula, valor predeterminado, etc. deben estar intactas.
- Las propiedades ACID de origen y destino deben validarse, etc.

### 3 - Verifique la carga de datos

(En un almacén de datos o en un mercado de datos o en cualquier lugar donde vaya a estar ubicado de forma permanente):

Los scripts reales que cargan los datos y los prueban definitivamente se incluirán en la prueba ETL. Sin embargo, el sistema de almacenamiento de datos debe validarse para lo siguiente:

- **Rendimiento:** A medida que los sistemas se vuelven más intrincados, se forman relaciones entre múltiples entidades para establecer varias correlaciones. Esta es una gran noticia para el análisis de datos, sin embargo, este tipo de complejidad a menudo hace que las consultas tarden demasiado en recuperar los resultados. Por lo tanto, pruebas de rendimiento juegan un papel importante aquí.
- **Escalabilidad:** Los datos solo aumentarán, no disminuirán. Por lo tanto, se deben realizar pruebas para asegurarse de que el tamaño del negocio en crecimiento y los volúmenes de datos puedan ser manejados por la implementación actual o no. Esto también incluye probar la estrategia de archivo. Básicamente, está tratando de probar la decisión: '¿Qué sucede con los datos más antiguos y si los necesito?'

También es una buena idea probar los otros aspectos como sus habilidades computacionales, recuperación de fallos, registro de errores, manejo de excepciones, etc.

#### 4 - Prueba de informe de BI

Finalmente, los informes, la última capa de todo el flujo.

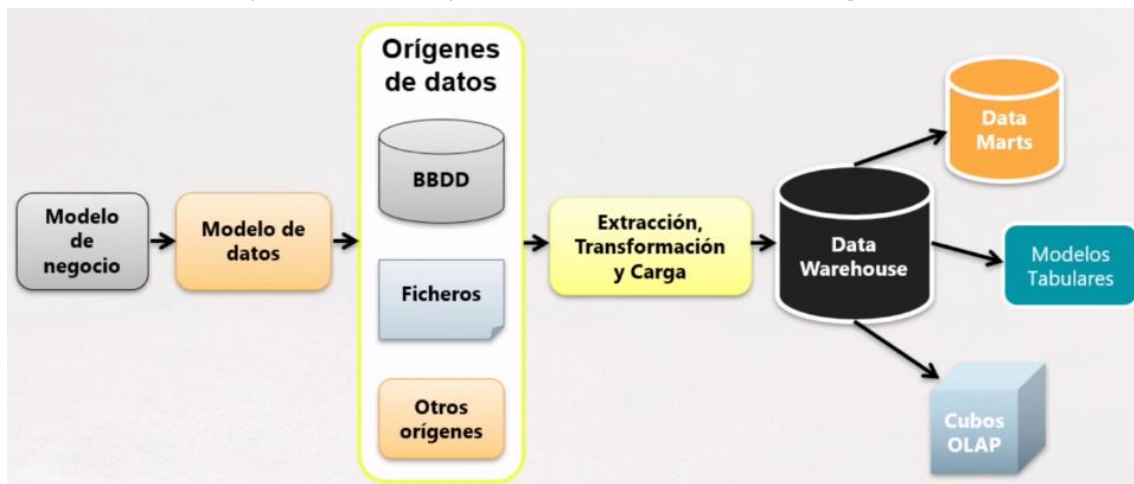
Esto es lo que se considera Business Intelligence. Pero, como puede ver en lo anterior, los informes nunca serán correctos, consistentes y rápidos si sus capas anteriores no funcionaban correctamente.

En este punto, debemos fijarnos en:

- Los informes generados y su aplicabilidad al negocio
- Posibilidad de personalizar y personalizar los parámetros que se incluirán en los informes. Ordenar, categorizar, agrupar, etc.
- Presentación del informe en sí.

## 5. Arquitectura típica del Business Intelligence

### 5.1. Arquitectura típica del Business Intelligence



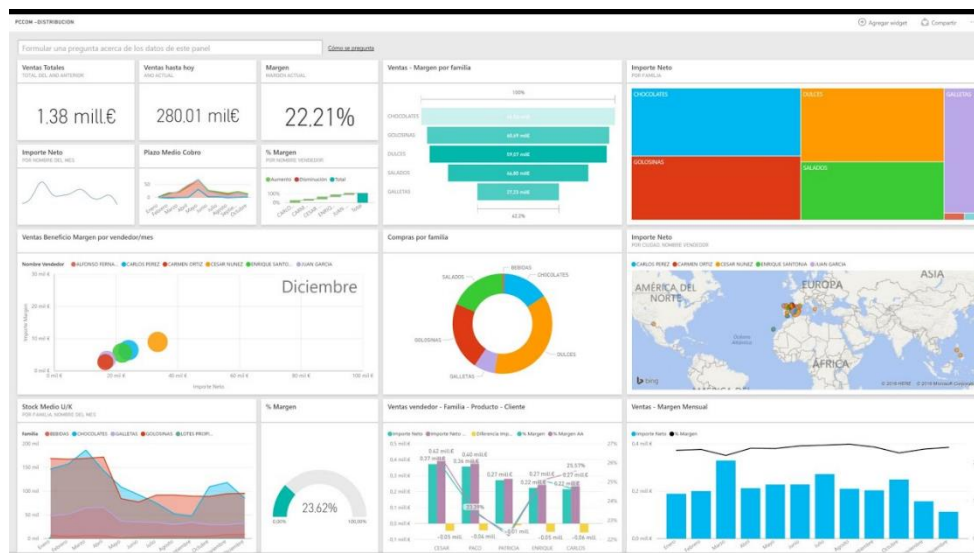
El esquema anterior representa un sistema BI típico, aunque puede sufrir variaciones.

- **Modelo de negocio:** son los datos generados en el proceso productivo de la empresa y que no están relacionados con el BI, es decir, el control que se realiza de un tanque con un líquido es parte del proceso y funciona sin BI.
- **Modelo de datos:** es la transformación de los datos generados por el proceso productivo y trasladados a un modelo donde los datos puedan ser almacenados y representados.
- **Orígenes de datos:** hace referencia a la fuente del que podemos extraer la información del sistema, por ejemplo, un PCL de una planta.
- **Extracción, transformación y carga:** con los datos almacenados, usando BI usaremos los datos, los transformaremos y los enviaremos al lugar donde realizaremos los tratamientos de análisis.
- **Data warehouse:** es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

## 5.2. Cuadros de mando

El análisis y la gestión de la información es una pieza clave en la transformación digital, un proceso estratégico para el futuro de todas las organizaciones. Sin embargo, en muchas ocasiones, los trabajadores no cuentan con suficiente tiempo, ni herramientas que les ayuden a toma de decisiones ágiles, lo que provoca que nuestras decisiones no se tomen en base a todos los datos disponibles.

Un cuadro de mando o dashboard es un panel que expone gráficamente los KPI's (Key Performance Indicator), conocido también como indicador clave de rendimiento, es una medida del nivel del rendimiento de un proceso, más importantes para el funcionamiento de un negocio. Permite el seguimiento, el análisis y la optimización de la estrategia de la empresa con el objetivo de facilitar la toma de decisiones.



### 5.2.1. Características de un dashboard

Para poder desarrollar adecuadamente un dashboard te aconsejamos que tengas en cuenta los siguientes aspectos:

- **Visual:** La información debe estar presentada de forma gráfica, limpia y ordenada para que nos permita entender intuitivamente los datos que estamos manejando. Del mismo modo es importante determinar un único objetivo por dashboard para focalizarte.

Intentar cubrir todos los objetivos en un solo panel lo sobrecargará de información, y no estarás atendiendo correctamente a ninguno.

- **Todo en una única hoja:** No sobrecargues el cuadro de mando de información, una sola hoja es suficiente para visualizar los datos más importantes. Además, facilita la presentación de los mismos y ayuda a la comprensión por parte de los usuarios.
- **Personalizado:** Un buen software de dashboard debe ser flexible y completo, para que se adapte a todas las empresas y estrategias. Además, cada usuario debe tener su propio cuadro de mando, ya que los datos y la profundidad de los informes varían según el destinatario.
- **Accionable:** Debe permitir hacer comparaciones, visualizar y contextualizar los datos con el fin de establecer valoraciones útiles.

### Consejos para definir un cuadro de mandos

- **Planificarlo antes de desarrollarlo:** Un buen cuadro de mando necesita una planificación. Estudia qué cantidad de información queremos medir y qué gráficos harán que visualices toda la información necesaria para medir la consecución de tus objetivos.
- **Contextualiza la información:** ¿Cómo saber si un dato es bueno, malo, poco, mucho... si no tenemos con qué contrastarlo? Incluye en tu panel gráficos históricos, rankings, benchmarks... con el objetivo de tener referencias a la hora de evaluar los datos.
- **Incluye KPIs optimizados:** Piensa en qué aspectos debes preocuparte para que tu negocio avance. Hay que elegir aquellos que aporten valor y tengan sentido para facilitar la toma de decisiones y alcanzar nuestros objetivos marcados.
- **Define tus objetivos:** Estos objetivos a corto plazo deben tener las siguientes características: específicos, medibles, alcanzables, realistas, y acotados en el tiempo.

### 5.3. Arquitectura BI: estructura tiempo real

Un sistema de tiempo de real es "**cualquier actividad de proceso de información o sistema que tiene que responder a estímulos generados externamente dentro de un plazo especificado y finito**". Consecuentemente, la correctitud de un sistema de tiempo real depende no sólo del resultado lógico de la computación, sino también del **tiempo en el que este resultado tarda en generarse**.

En los sistemas de tiempo real el computador actúa generalmente junto con dispositivos físicos. Está dedicado al control y la monitorización de estos equipos y forma parte de un sistema de ingeniería más amplio, como por ejemplo una turbina en una central eléctrica. Por esta razón, los sistemas de tiempo real se denominan así mismos **sistemas empotrados**.

Un sistema en tiempo real se compone de:

- **Componente de adquisición de datos:** adquiere y formatea los datos procedentes del exterior.
- **Componente de análisis:** transforma los datos recogidos según lo requiera la aplicación.
- **Componente de control/salida:** genera la respuesta necesaria en cada momento.
- **Componente de monitorización:** coordina todos los componentes.

#### 5.3.1. Características de los sistemas en tiempo real

**Grandes y complejos:** puesto que los STR deben responder a los eventos del mundo real, la variedad de los mismos conduce a aplicaciones de gran tamaño. Por ello, los lenguajes de tiempo real proporcionan facilidades para poder dividirlos.



**Determinismo:** es una cualidad clave en los sistemas de tiempo real. Es la capacidad de determinar con una alta probabilidad, cuanto es el tiempo que se toma una tarea en iniciarse. Esto es importante porque los sistemas de tiempo real necesitan que ciertas tareas se ejecuten antes de que otras puedan iniciar.

**Responsividad:** La responsividad se enfoca en el tiempo que tarda una tarea en ejecutarse una vez que la interrupción ha sido atendida. Los aspectos a los que se enfoca son:

- La cantidad de tiempo que se lleva el iniciar la ejecución de una interrupción
- La cantidad de tiempo que se necesita para realizar la tarea que pidió la interrupción.
- Los efectos de interrupciones anidadas.

**Fiabilidad y seguridad:** El sistema no debe solamente estar libre de fallas, sino que además, la calidad del servicio que presta no debe degradarse más allá de un límite determinado. El sistema debe de seguir en funcionamiento a pesar de catástrofes, o fallas mecánicas. Usualmente una degradación en el servicio en un sistema de tiempo real lleva consecuencias catastróficas.

**Interacción con el hardware:** la naturaleza de estos sistemas empotrados es la conexión con elementos del mundo real como sensores o actuadores. Hoy en día, debido a su carácter crítico, el control se realiza directamente, sin la supervisión del S.O., por tanto, los lenguajes de tiempo real deben proporcionar primitivas de acceso a los dispositivos y soporte a interrupciones.

## Rendimiento

Cada diseño de tiempo real relativo al software debe ser aplicado en el contexto del “rendimiento” del sistema. Puesto que **el rendimiento** de un sistema de tiempo real se **determina** principalmente por el **tiempo de respuesta del sistema** y su **razón de transferencia** de datos, es importante comprender estos dos parámetros.

- El **tiempo de respuesta** del sistema es el tiempo en el que un sistema debe detectar un suceso interno o externo y responder con una acción. El procesamiento de la información sobre el suceso para determinar la respuesta adecuada puede implicar algoritmos complejos que consumen mucho tiempo. Entre los parámetros clave que afectan al tiempo de respuesta:
  - **Cambio de contexto:** se refiere al tiempo y sobrecarga necesitado para conmutar entre tareas.
  - **Latencia de interrupción:** es el tiempo que pasa antes de que el cambio sea realmente posible.
  - **Otros factores:** otros parámetros que afectan al tiempo de respuesta son la velocidad de cálculo y el acceso a memorias masivas.
- **La razón de transferencia de datos** es la velocidad con que los datos entran o salen del sistema.

Aunque todas las aplicaciones de software deben ser fiables, los sistemas de tiempo real hacen una especial demanda de fiabilidad, reinicialización y recuperación de fallos. Debido a que el mundo real está siendo monitorizado y controlado, la pérdida de monitorización o control (o ambos) es intolerable en muchas circunstancias (p. ej.: un sistema de control de tráfico aéreo). Consecuentemente, los sistemas de tiempo real contienen mecanismos de restauración y recuperación de fallos y, frecuentemente, tienen incorporadas redundancias para asegurar la restauración.

## Bases de datos en tiempo real

Los sistemas de tiempo real requieren normalmente del uso y gestión de una base de datos. Estas bddd suelen ser **bases de datos distribuidas**, debido a que la multitarea es muy común y que los datos se procesan frecuentemente en paralelo. Si la base de datos es distribuida y no centralizada, las tareas individuales pueden acceder a sus datos de forma más rápida, fiable y con menos cuellos de botella.

El uso de una base de datos distribuida para aplicaciones de tiempo real divide el tráfico de entrada/salida y acorta las colas de las tareas en espera, para acceder a una base de datos. Además, un fallo de una base de datos raramente causará el fallo del sistema entero si se construyen con redundancia.

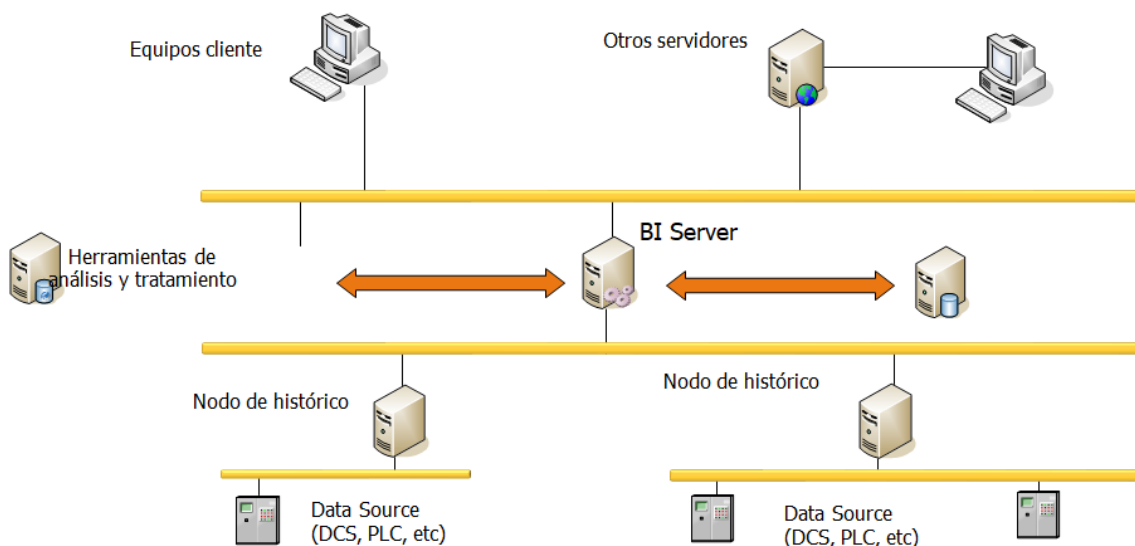
Uno de los elementos a tener en cuenta con una bddd de tiempo real, es el almacenamiento de datos, ya que algunos de ellos pueden venir de sensores que muestrean en períodos muy cortos de tiempo, lo cual nos hace preguntarnos, ¿es necesario almacenar toda la información? Un buen sistema en tiempo real ha de ser capaz de almacenar la información relevante en la bddd, de esta forma se ahorra espacio.

Ejemplo:

- SAP HANA
- PI System

## 5.4. Ejemplo BI industrial

Los sistemas BI sirven para empresas de diferentes ámbitos, desde una tienda de comercio online hasta una planta de producción de plásticos. En el siguiente ejemplo se puede apreciar un ejemplo de entorno BI industrial, el cual recoge de diferentes fuentes de datos, la información muestreada por los dispositivos de captación.



La información llega desde los orígenes de datos hasta un servidor intermedio, dónde se almacenan los datos en crudo (RAW data), de ahí, un servidor de almacenamiento con un sistema BI, recoge los datos, los selecciona y los almacena para futuras consultas.

El tipo de sistemas que aparece en la imagen anterior suele relacionarse con un sistema de tiempo real, los cuales realizan el muestreo de datos en períodos muy cortos de tiempo. Esto conlleva una serie de problemas que los administradores del sistema deben tener en cuenta:

- Al muestrear datos en períodos cortos de tiempo, habrá información redundante
- Se debe almacenar información relevante para los procesos
- El almacenamiento, a pesar de ser informático, no es infinito

Debido a los problemas expuestos, se necesita un método para poder almacenar los datos de manera eficiente y a la vez no perder información relevante.

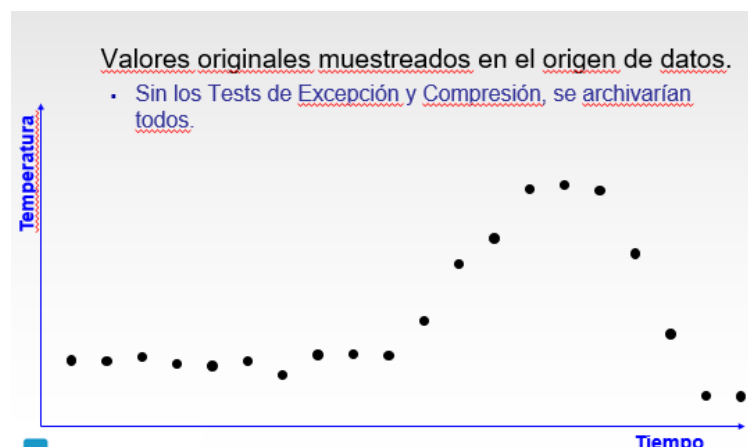
Algunos softwares de captación de tiempo real, realizan varios test a los datos en crudo, de forma que solo se almacenan los datos relevantes. Podemos hablar, por ejemplo, de los siguientes tests:

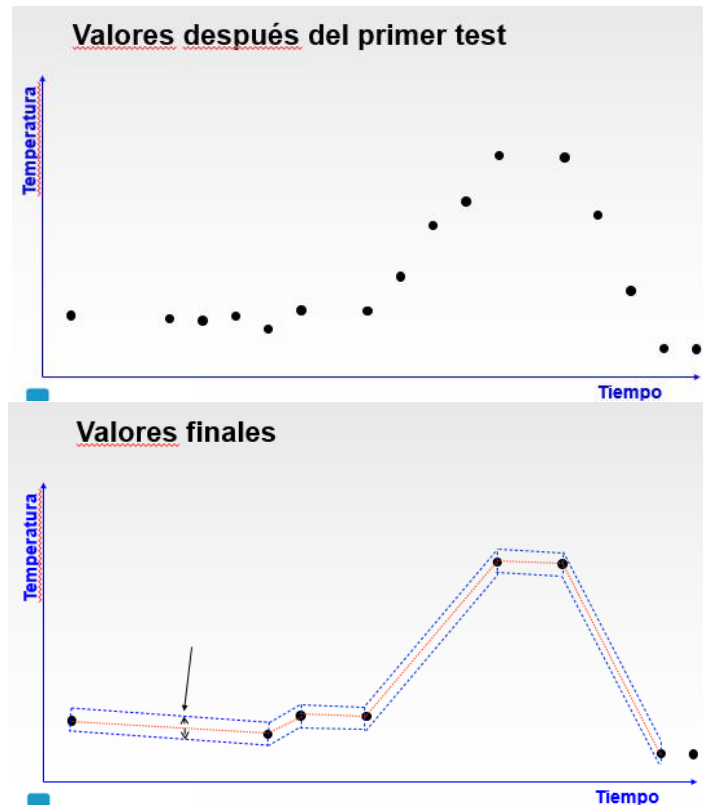
- **Test de valor:** en función de un valor, se determina si el siguiente es relevante o no.
- **Test de tiempo:** en función del tiempo, se almacena o no el dato

**Test de valor:** se define un margen inferior y superior en base al último valor recibido desde el origen de datos. Si un valor supera ese margen, se almacena dicho valor y su anterior, así recogerá un histórico de cómo ha ido evolucionando la señal, por ejemplo, un 10%. Ejemplo:

Valor inicial: 12  $\rightarrow$  10% = 1,2  
 Nuevo valor <sub>1</sub>: 11,5  $\rightarrow$  no se almacena  
 Nuevo valor <sub>2</sub>: 10,1  $\rightarrow$  se almacena  
 Valor inicial: 10,1  $\rightarrow$  10% = 1,01  
 Nuevo valor <sub>3</sub>: 13,1  $\rightarrow$  se almacena

**Test de tiempo:** el test anterior quita muchos valores repetidos, pero puede darse el caso que durante un período prolongado, el dato no supere el test de valor y se almacene, por lo que para corregir esto, se define un tiempo máximo en el que el sistema puede estar sin almacenar datos, por ejemplo 40 segundos. Si pasado ese tiempo, no se ha cumplido el test de valor, se almacena el último valor, pasando a ser éste el nuevo valor inicial





## 6. Bibliografía

- <https://www.threepoints.com/es/3-ejemplos-de-business-intelligence>
- <https://notjustbi.com/modelos-de-datos-en-bi-business-intelligence/>
- <https://www.javatpoint.com/kdd-process-in-data-mining>
- <https://www.campusmvp.es/recursos/post/el-proceso-de-extraccion-de-conocimiento-a-partir-de-bases-de-datos.aspx>
- <https://blogvisionarios.com/articulos-data/testing-en-business-analytics/>
- <https://spa.myservername.com/4-steps-business-intelligence-testing>
- Jiménez, L.M.; Puerto, R. (2017). Sistemas informáticos en tiempo real: teoría y aplicaciones. Amazon Media.
- López Herrera, JF. (2011). Programación en tiempo real y bases de datos: un enfoque práctico. Universidad politécnica de Cataluña.
- Pastor, F. (2005). Sistemas informáticos en tiempo real. Universidad de Valencia