

# Business Intelligence



red.es



*"El FSE invierte en tu futuro"*  
**Fondo Social Europeo**

# Índice

## 1. Datos y decisiones de negocio

- 1.1 Qué es Business Intelligence
- 1.2 Del dato a la decisión
- 1.3 Estructura BI
- 1.4 BO y BI
- 1.5 Ejemplos BI

## 2. Modelos de datos

- 2.1 Data Warehouse
- 2.2 Data mart
- 2.3 OTLP
- 2.4 OLAP
- 2.5 Modelo estrella
- 2.6 Modelo copo de nieve
- 2.7 Drilldown, drillup y drillthrough
- 2.8 ETL

## 3. Soluciones BI

- 3.1 Vista atrás
- 3.2 Tipo de salidas
- 3.3 Beneficios de BI
- 3.4 Herramientas de BI

## 4. Minería de datos

- 4.1 Qué es
- 4.2 Ramas de la minería de datos
- 4.3 Árbol de decisión
- 4.4 Clusters
- 4.5 Redes neuronales

## 5. Ejercicios

# 1. Datos y decisiones de negocio



## 1.2 Del dato a la decisión

El BI implica varios términos a tener en cuenta:

- Decisiones
- Personas
- Información

**BUENAS DECISIONES = INTELIGENCIA + INFORMACIÓN**

MALAS DECISIONES = **INTELIGENCIA** + INFORMACIÓN

MALAS DECISIONES = INTELIGENCIA + **INFORMACIÓN**



## 1.3 Estructura BI

En la estructura de Business Intelligence participan:

### 1. Datos

Los cuales por sí solos no aportan ninguna información y están dispersos por toda la organización y en distintos formatos

### 2. Información

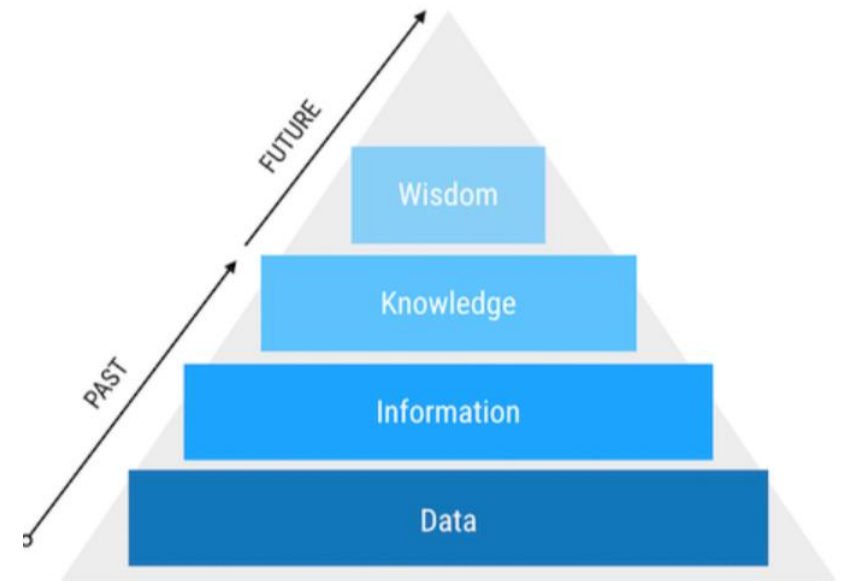
Reunión de todos los datos en un formato en el que pueda leerlo.

### 3. Conocimiento

Se deriva de las personas y es intangible y empírico.

### 4. Decisiones

Implica el funcionamiento de un sistema de BI implementado que me permite tomar decisiones.



## 1.4 Estructura BI

### ¿Qué entendemos por información?

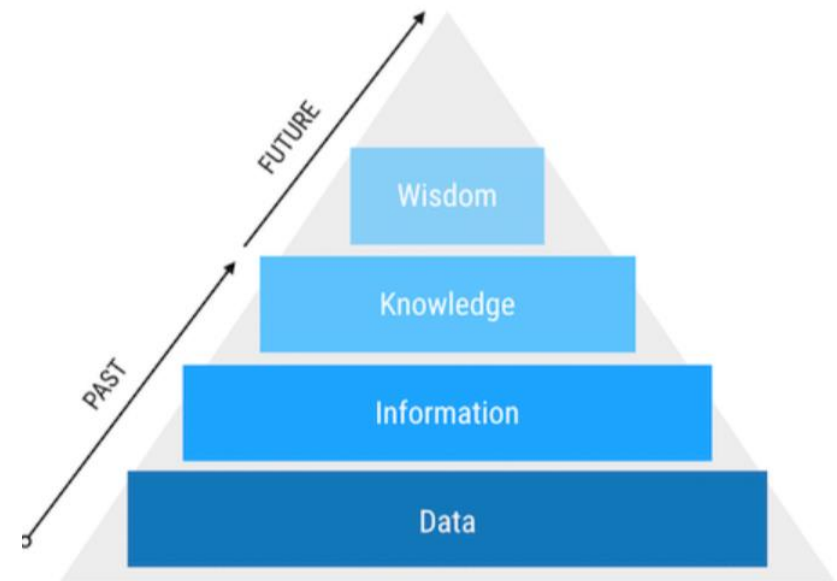
Toda la información que tengamos identificada, categorizada, etiquetada o calculada tras la recogida de datos.

#### **BO:** Business Operation

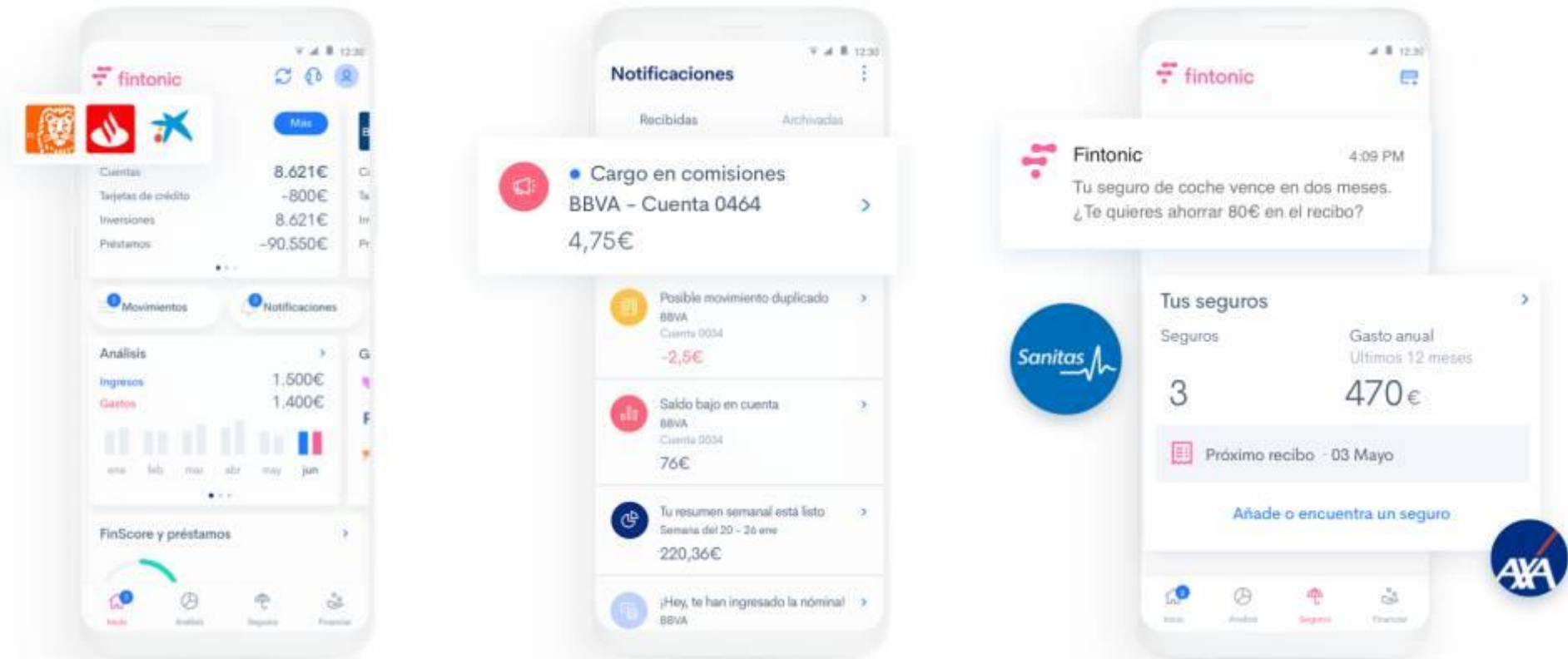
- Estructura y orden básico aplicado a los datos
- Tablas, documentos, listas, carpetas, etc.

#### **BI:** Business Intelligence

- Estructura y orden aplicada a los datos
- *Data Warehouse, DataMart*

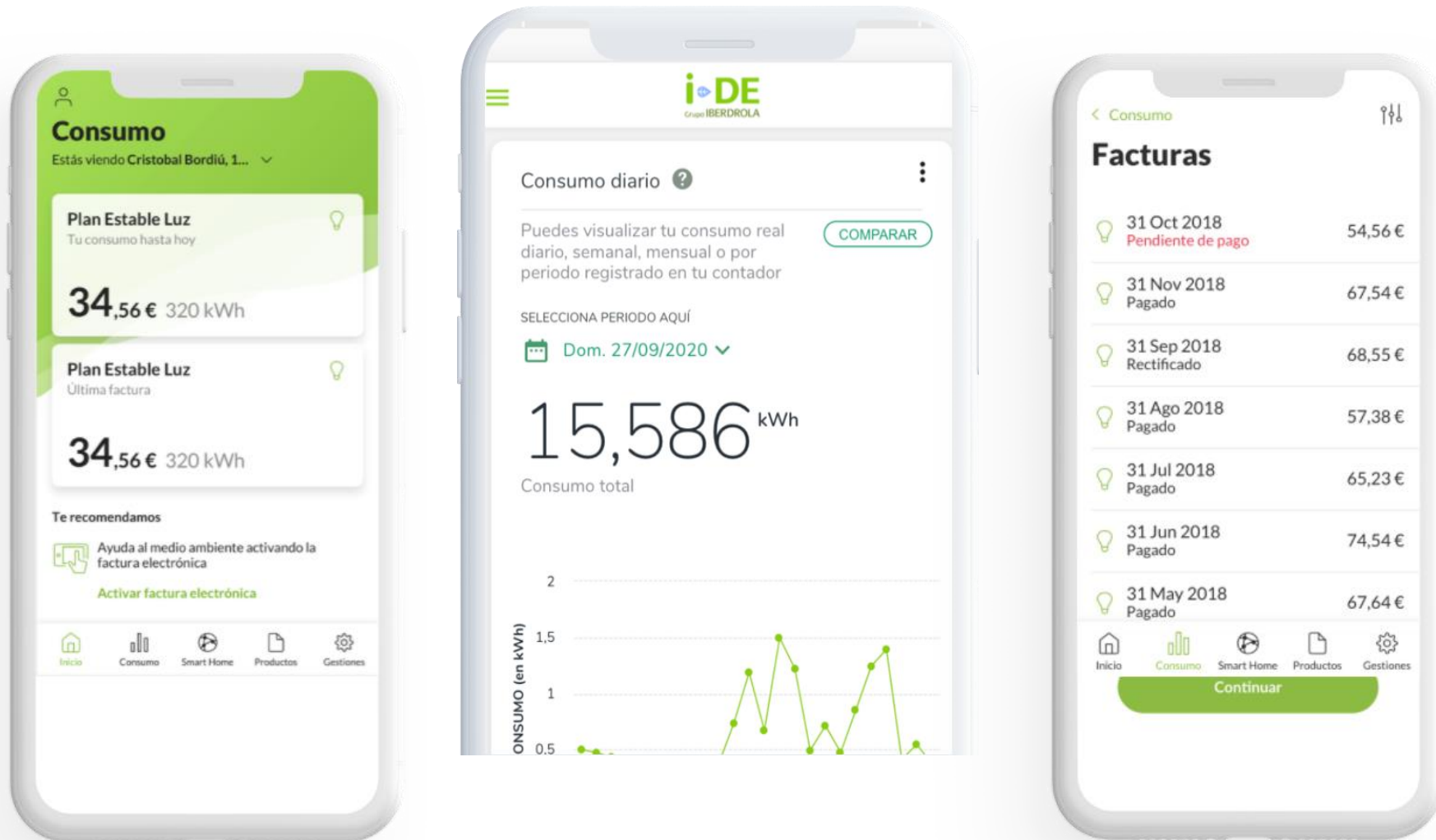


## 1.5 Ejemplos BI





## 1.5 Ejemplos BI



## 2. Modelos de datos

## 2.1 El Data Warehouse

En la mayoría de ocasiones, el *data warehouse* será el primer paso desde el punto de vista técnico para implantar una solución completa y fiable de BI.

Al organizar los datos en una ubicación, tus empleados pueden resolver problemas más rápido y cumplir con los plazos de manera constante.

Con un *data warehouse* la información fluye de manera constante mientras los analistas la revisan.

- **Orientado a temas:** Suelen tratar información similar y está unida entre sí. Por ejemplo: todas las transacciones bancarias en las que se incluya el importe, el origen, la fecha y hora, etc.
- **No es variable en el tiempo:** Almacena un histórico de datos
- **No es volátil:** No se elimina ni modifica la información reciente. Esta información se mantiene intacta y en su estado original.
- **Integrado:** Permite integrar distintas fuentes de datos (por ejemplo, áreas de la compañía) que se van a poder exportar a un mismo lugar.

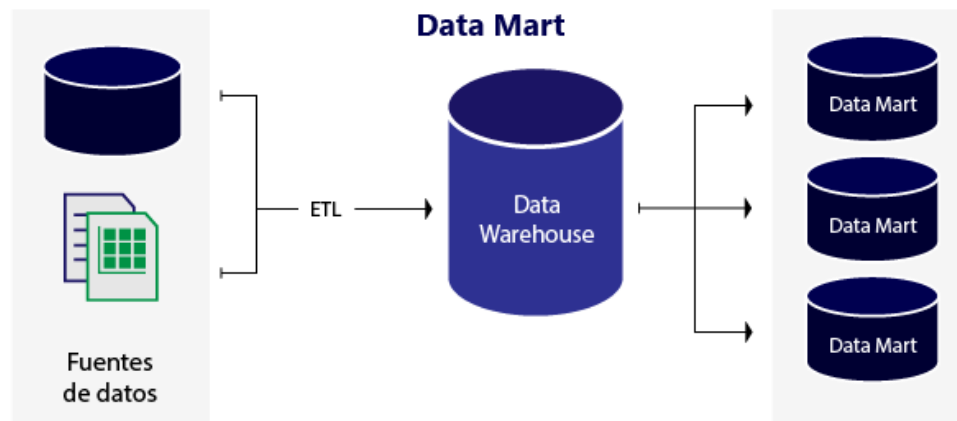
## 2.2 El Data Mart

Son almacenes de datos con información de interés particular para un determinado sector de la empresa.

Por ejemplo, podemos tener un data mart para estas áreas:

- **Recursos Humanos**
- **Operaciones**
- **Ventas**

**Data Warehouse** = DM de Ventas + DM de Recursos Humanos + DM de Producción



## 2.3 OTLP

¿Por qué necesitamos un Data Warehouse? ¿No podemos seguir trabajando con bases de datos de la forma tradicional?

Hasta ahora hemos trabajado con **bases de datos transaccionales**, que reúnen las siguientes características:

- Optimizadas para transacciones cortas del tipo (INSERT, UPDATE, DELETE)
- Para análisis, las estructuras tradicionales no son ágiles
- Los datos no están organizados para ser agrupados mediante relaciones que nos permitan dotarlos de significado. No tenemos información.
- Suelen estar en formato 3FN

STUD_NO	STUD_NAME	STUD_STATE	STUD_COUNTRY	STUD_AGE
1	RAM	HARYANA	INDIA	20
2	RAM	PUNJAB	INDIA	19
3	SURESH	PUNJAB	INDIA	21

## 2.4 OLAP

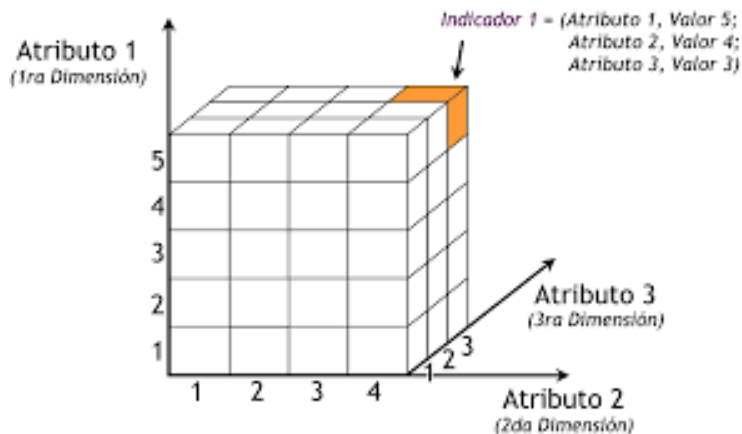
OLAP es una tecnología para la inteligencia de negocios que ejecuta análisis multidimensional de datos de una forma veloz e interactiva.

No está optimizado para transacciones. Implican cargas pesadas en procesos (normalmente nocturnos) que involucran agregados.

Analizan las relaciones entre muchos tipos de datos y/o elementos empresariales.

Utiliza esquemas de tipo 'Estrella' o 'Copo de nieve'

**¿Cuánto dinero he ganado vendiendo el producto X en la región Y en el periodo de tiempo Z?**



**[CLICA PARA VER VÍDEO OLAP](#)**

## 2.4 OLAP

De OLAP surgen varios conceptos que vamos a ver de ahora en adelante:

### Hechos

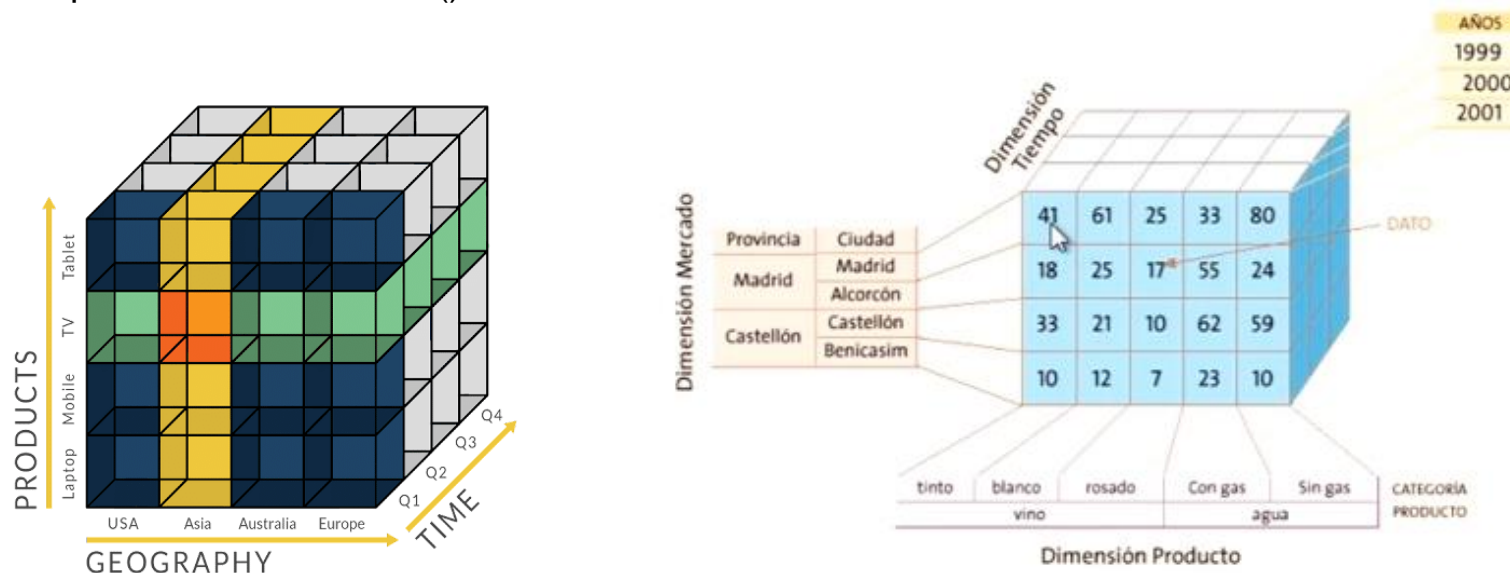
Por ejemplo, el salario o la edad de un empleado.

### Dimensiones

Son propiedades de tipo cualitativo. Por ejemplo la región o el tiempo.

### Agregaciones

Los hechos pueden agregarse para facilitar cálculos a la herramienta BI. Por ejemplo, que el número de empleados sea una SUM() de todos los departamentos.



## 2.5 Esquema de estrella

Un *esquema de estrella* es un tipo de esquema de base de datos relacional que consta de una sola tabla de hechos central rodeada de tablas de dimensiones.

Las **tablas de hechos** pueden almacenar observaciones o eventos, y pueden ser pedidos de ventas, existencias, tasas de cambio, temperaturas, etc. Una tabla de hechos contiene columnas de clave de dimensiones relacionadas con las tablas de dimensiones y columnas de medida numéricas.

Esquema de estrella

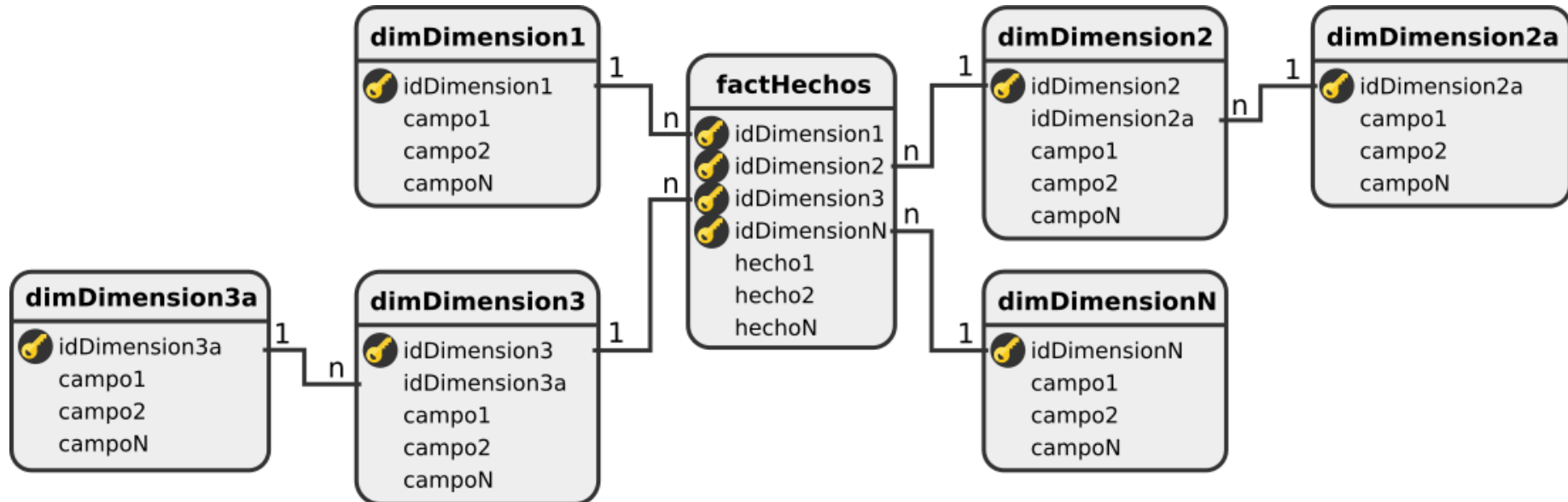




## 2.6 Esquema de copo de nieve

El *esquema de copo de nieve* consta de una tabla de hechos que está conectada a muchas tablas de dimensiones, que pueden estar conectadas a otras tablas de dimensiones a través de una relación de muchos a uno.

Las tablas de un esquema de copo de nieve generalmente se normalizan en el tercer formulario de normalización. Cada tabla de dimensiones representa exactamente un nivel en una jerarquía.



## 2.7 Jerarquía: Drill-down, roll-up y drill-through

### **Drill down:**

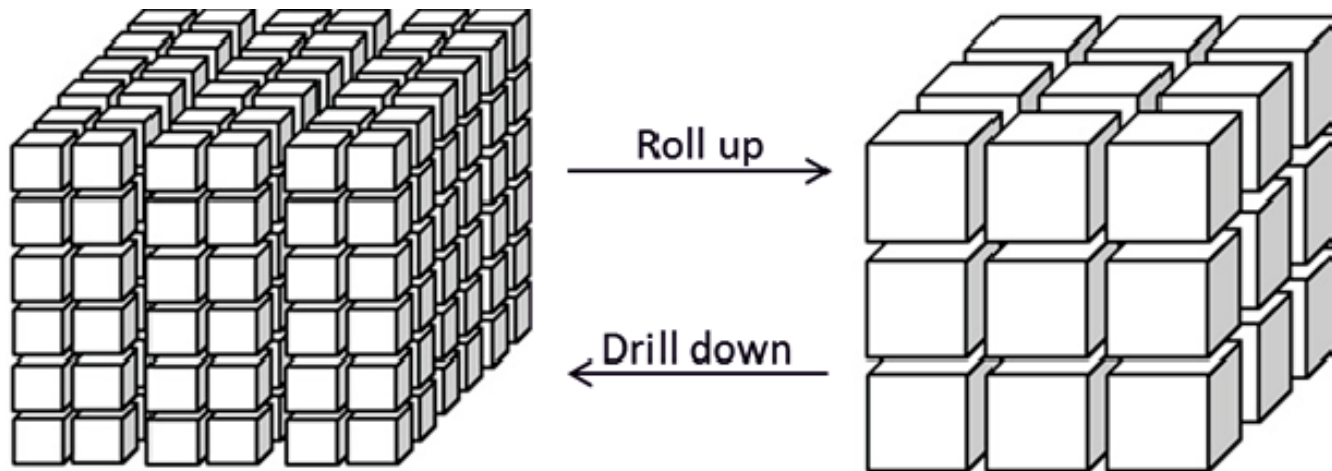
Navegamos de un nivel superior de una información a uno inferior para aumentar el nivel de granularidad de los datos.

### **Roll Up:**

Partimos desde el nivel más bajo a uno más alto para recoger información

### **Drill Through:**

Navegamos de forma horizontal dentro de un mismo nivel de información.



## 2.8 ETL

ETL significa *extracción, transformación y carga*. Es un método de integración de datos que no están optimizados que consiste en extraer, transformar y cargar múltiples fuentes de información para almacenarlas en un solo destino o almacén de datos que simplifica su gestión y análisis.

### EXTRACCIÓN

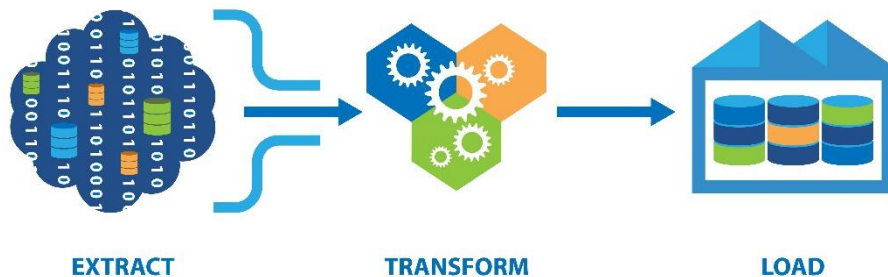
Obtención de la información de las distintas fuentes tanto internas como externas. La velocidad y el orden de extracción de dicha información tienen un gran impacto en todo el proceso de integración.

### TRANSFORMACIÓN

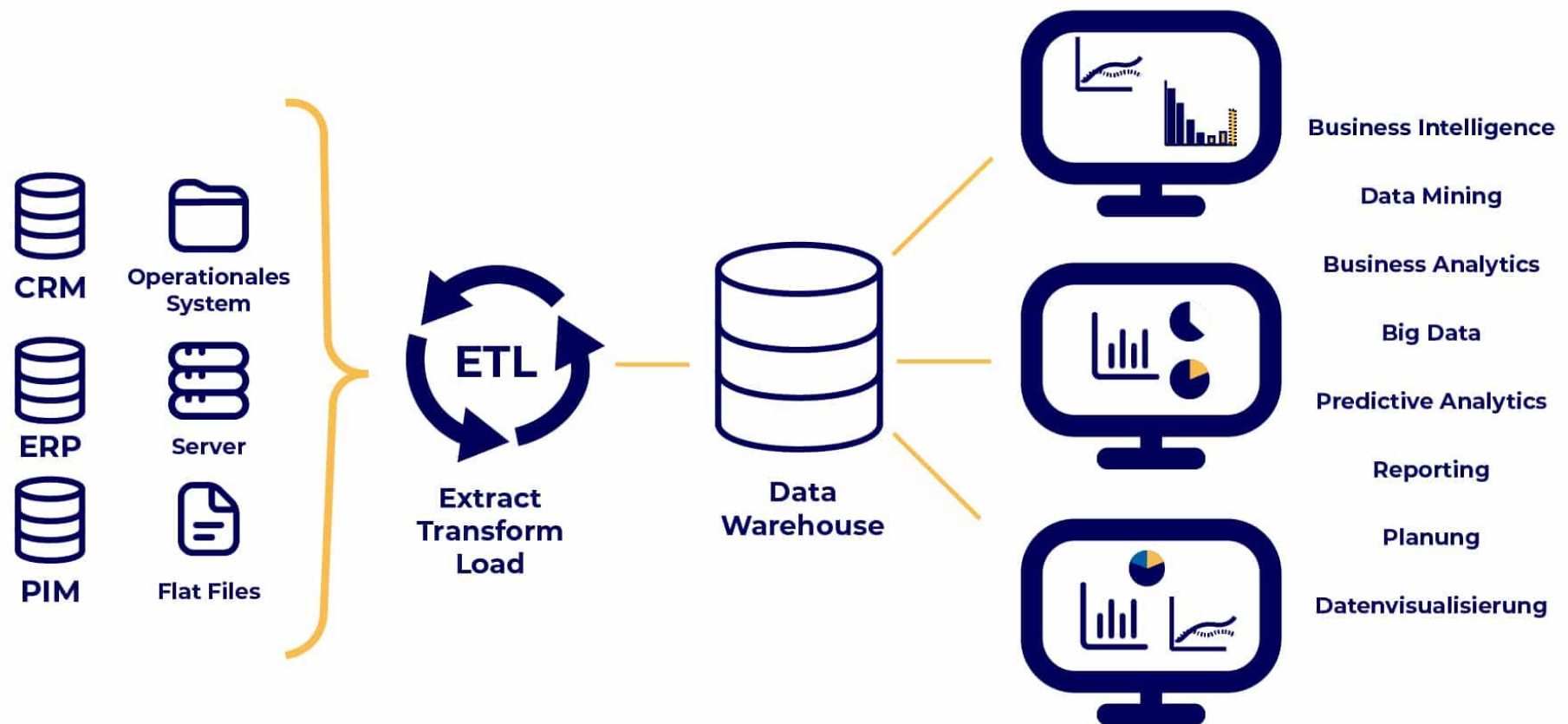
Filtrado, limpieza, depuración, homogeneización y agrupación de la información.

### CARGA

Organización y actualización de los datos y los metadatos en el DW.



## 2.8 ETL



# 3. Soluciones BI

## 3.1 Vista atrás

¿No se hacían informes inteligencia de negocio antes de la llegada de soluciones BI en las empresas?

¿Cómo eran estos informes?

### **Informes Estáticos**

Los informes eran poco flexibles y no tenían capacidad de cambiarlos o manipularlos ágilmente.

### **Dependencia del área de IT**

Había muchas peticiones y falta de eficiencia por la alta dependencia de perfiles de carácter tecnológico para poder obtener los datos deseados, efectuar los cambios o modificaciones correspondientes y así tomar decisiones de negocio.

### **Falta de entendimiento**

Los distintos departamentos o áreas de las empresas no hablaban el mismo idioma, dado que cada uno utilizaba informes y herramientas distintas.

## 3.2 Salidas BI

Las herramientas de BI nacen para dotarnos de mayor flexibilidad y homogeneidad ante la manera antigua de construir esos informes y tomar decisiones. En este apartado, encontramos lo que conocemos como tipos de salidas en BI:

### DSS

Son los llamados '**Sistemas de soporte a la decisión**'. Comprenden informes dinámicos y no requieren conocimientos técnicos. La información está dirigida y adecuada a cada perfil.

### EIS

Son los llamados '**Sistemas de información ejecutiva**'. Ofrecen indicadores de negocio o KPI y permiten análisis de expectativas y por supuesto, apoyan la toma de decisiones empresariales.

### CMI

También llamados "**Cuadro de mando integrales**". Orientados a la toma de decisiones por altos puestos directivos y agrupan todos los departamentos de la compañía.



## 3.3 Beneficios BI

### **Seguimiento real**

Será posible detectar a tiempo las decisiones del negocio conforme a la estrategia predefinida, adoptando las acciones oportunas para corregirlas.

### **Aprender de errores pasados**

Podemos acceder a análisis de datos históricos para detectar errores pero también nuevas oportunidades en la toma de decisiones

### **Mejorar la competitividad**

A partir del uso de un sistema BI, toda la información disponible y relevante de la empresa va a empujar en la dirección única de mejorar las decisiones tomadas.

### **Abstracción de la tecnología**

La toma de decisiones por parte de cualquier persona que tenga que hacerlo no debe depender de que los usuarios conozcan la tecnología subyacente.





## 3.4 Herramientas BI

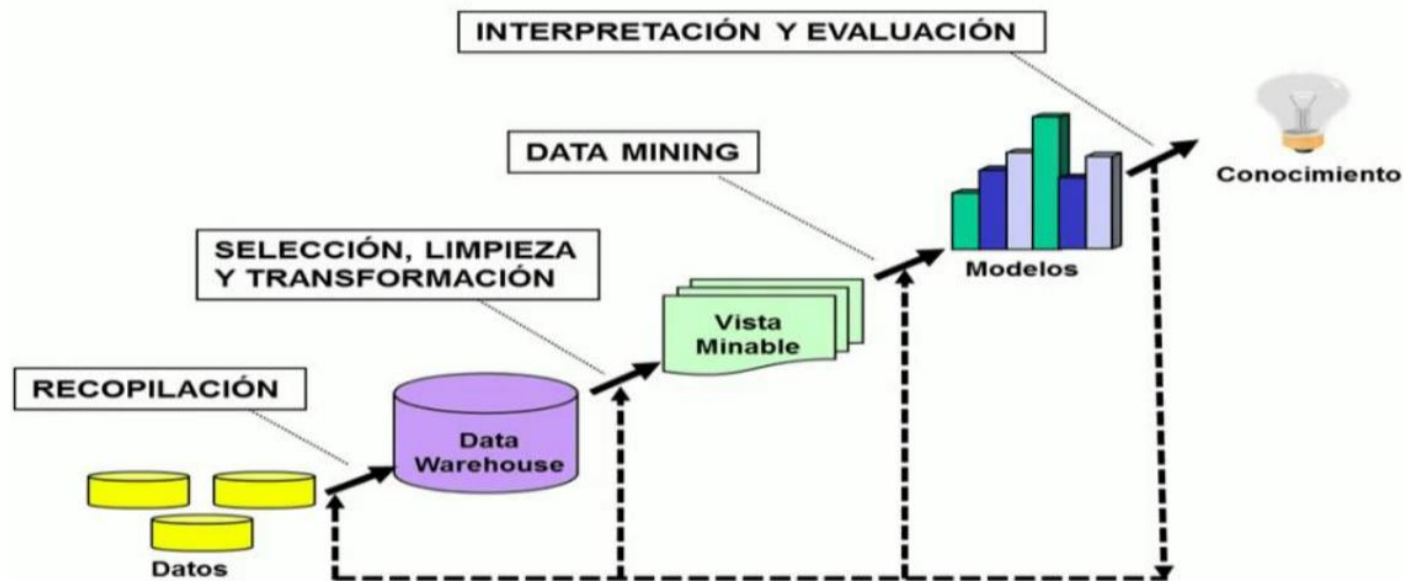


# 4. Minería de datos

## 4.1 Qué es

La minería de datos es una rama más ligada a la estadística y la matemática que a la informática.

- Muchas herramientas de BI ya integran soluciones de minería de datos.
- Podemos sacar información útil que simplemente con una análisis de BI no hemos conseguido sacar
- Permiten y son capaces de predecir tendencias de hechos futuros.



## 4.2 Ramas

A grandes rasgos, podríamos dividir la minería de datos en dos ramas:

### Estadística clásica

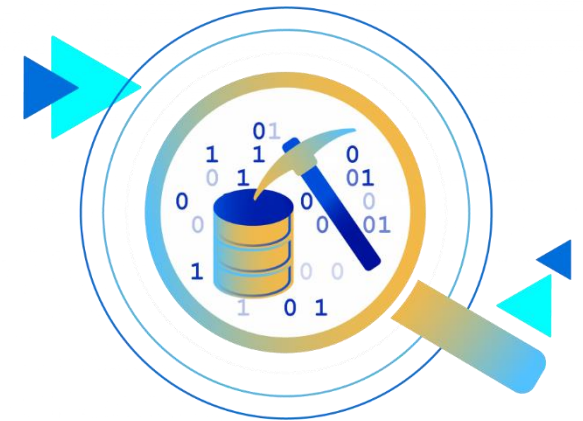
Se utiliza principalmente con un fin puramente predictivo y para ello podemos hacer uso de:

- ✓ Árboles de decisión
- ✓ Clustering
- ✓ Análisis de regresión
- ✓ Etc.

### Minería de datos actual

Está basada en inteligencia artificial y aprendizaje automático y además de predecir, se usa para descubrir conocimiento.

- ✓ Redes neuronales
- ✓ Agrupamiento k-means
- ✓ Etc.

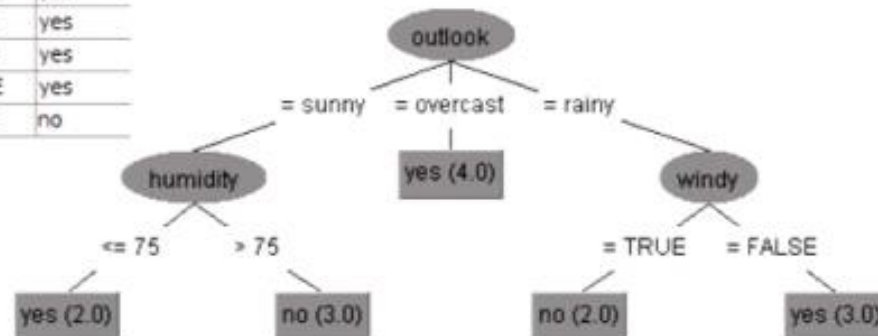


## 4.3 Árbol de decisión

Un árbol de decisión es un mapa de los posibles resultados de una serie de decisiones relacionadas. Permite que un individuo o una organización comparen posibles acciones entre sí según sus costos, probabilidades y beneficios.

Se pueden usar para dirigir un intercambio de ideas informal o trazar un algoritmo que anticipe matemáticamente la mejor opción.

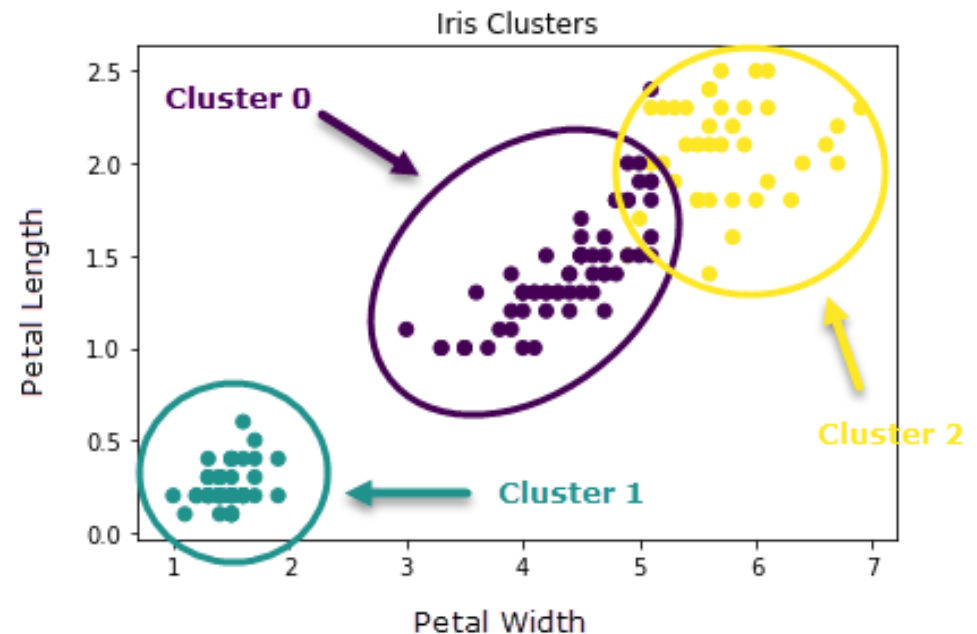
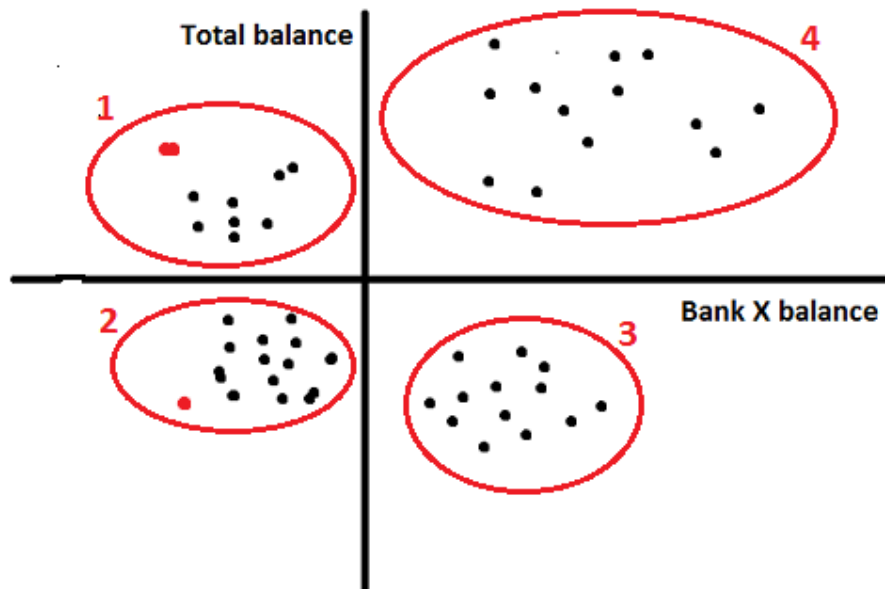
No.	outlook Nominal	temperature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	70.0	FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	72.0	90.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no



## 4.4 Clustering

Clustering es una técnica utilizada en minería de datos (dentro del área de la Inteligencia Artificial) para **identificar de forma automática agrupaciones** (clústeres) de elementos de acuerdo a una medida de similitud entre ellos. Esta técnica también se conoce como segmentación.

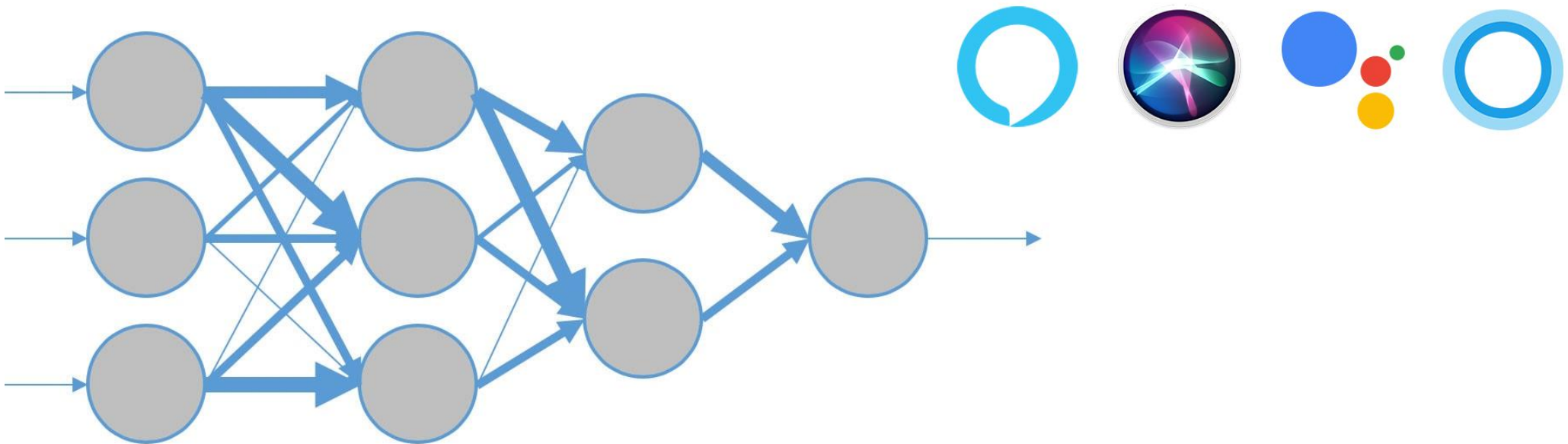
En el área del Business Intelligence, la técnica de clustering puede ser utilizada para organizar diferentes tipos de datos tales como productos, clientes o tiendas.



## 4.5 Redes neuronales

Las redes neuronales artificiales son un modelo computacional que permite **simular el comportamiento del cerebro humano**, es decir, dotar a las máquinas de la capacidad de aprender de una manera similar a como lo hace nuestro cerebro.

Una red neuronal artificial está formada por neuronas artificiales, que son unidades o nodos que **reciben información del exterior o de otras neuronas**, de manera similar a los impulsos nerviosos que reciben las neuronas del cerebro humano, las procesan y generan un valor de salida que alimenta a otras neuronas de la red o son la salida hacia el exterior de la red.



# 5. Ejercicios



## 5. Ejercicios prácticos

### E.1

Diseña en una hoja con papel y boli un cuadro de mando para el proyecto del restaurante de ayer.

### E.2

Descarga el dataset de Kaggle '**Netflix Movies and TV Shows**'. Cárgalo en Data Studio e intenta responder a las preguntas que plantea.

### E.3

Descarga el dataset que tú prefieras y monta un cuadro de mando integral en Data Studio.

**ANEXO:** [Vídeo de tablas cruzadas](#). LEFT JOIN, RIGHT JOIN, INNER JOIN-



red.es

Centro de  
Referencia Nacional  
en Comercio Electrónico  
y Marketing  
CRN  
Digital



*"El FSE invierte en tu futuro"*

**Fondo Social Europeo**

