



MASTER DATA SCIENCE

# Modelización Multidimensional I

Máster Data Science HINTD-EFBS

Javier Picos

05 de Noviembre de 2021



# Contenido

- Descripción del curso
- Plataformas de datos modernas
- Datawarehousing
- Fundamentos de BBDD Relacionales
- Modelización Multidimensional
- Proyecto de Implantación

- Ingeniero Industrial (Organización Industrial)
- Máster Logística CEL
- Certificado CPIM por APICS
- Certificado CPM por ISM
- Socio en HINTD

# Descripción del curso

Objetivos y  
Prerequisitos

# Descripción del curso

## Objetivos

- Un buen **modelo dimensional de datos**, o **esquema en estrella**, es la base de un excelente sistema de **Business Intelligence / Datawarehouse**.
- Se explicará por qué es importante dentro de una organización, disponer de una plataforma **Datawarehouse (independiente, escalable, sostenible y con rendimiento)**.
- En este curso aprenderán **técnicas prácticas para diseñar, desarrollar e implementar un Modelo Multidimensional consistente**.
- La aproximación se hará desde un conocimiento previo de las bases de datos relacionales de los sistemas operacionales y sus fundamentos, como contraposición a los Datawarehouse (basados en modelos multidimensionales) que son la base de los sistemas analíticos (business Intelligence entre otros).

## Conocimientos previos

- Se requiere un conocimiento básico de los **procesos de negocio de una organización típica empresarial (Comercial/Ventas, Compras, Producción, Económico-Financieros, Personas)**.
- Se requiere un conocimiento básico de los **sistemas de información transaccionales/operacionales (origen y fuente de datos)**.
- Se requiere un conocimiento básico de los **modelos de datos (Entidad-Relación), Integridad referencial y lenguaje SQL (Structured Query Language)**.

→ Descripción del curso

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

# Plataformas de datos modernas

Ventaja  
competitiva

# Plataformas de datos modernas | Transformación digital

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

Ecuación de valor

Líneas de acción para la Transformación Digital

Tecnologías facilitadoras

↑ Ingresos

↓ Costes

↓ Circulante

Organización

Liderar el cambio  
Adaptar la estructura organizativa  
Capacitar a las personas  
Gestionar talento

Plataforma Data Analytics

Arquitectura de datos  
Independencia infraestructura  
Rendimiento  
Configurable  
Escalable

Negocio

Eco-Fin  
Ventas y Marketing  
Clientes  
Operaciones  
Personas

Conectividad

IoT  
Cloud  
Ciber Seguridad

Apps Cliente

Apps Móviles  
RRSS  
Interacción con cliente

Automatización

Visión artificial  
Sensorización  
Impresión 3D  
Robótica

Data & Analytics

Big Data  
Analítica Avanzada  
(Predictiva y Prescriptiva)

# Plataformas de datos modernas | 3 puntos de vista

## → Plataformas de datos modernas

### → Datawarehousing

### → Bases de datos relacionales

### → Modelización Multidimensional

### → Proyecto de implantación

## Qué son?

## Características principales?

## Qué se ha de considerar?

### Tecnológico

- Las plataformas de Datos **son una amalgama de sistemas**: BBDD relacionales, No SQL, Almacenamiento Blob, Sistemas Analíticos,...

- Cada componente o elemento de la arquitectura tecnológica tiene una orientación específica** (sistemas de almacenamiento masivo, sistemas transaccionales, sistemas analíticos,...)

- Flexibilidad a la hora de **escalar**
- Seguridad** y regulación
- Capacidad de **virtualización**
- Independencia de infraestructura**
- Rendimiento**
- Facilidad de mantenimiento, operación y uso**

### Datos

- Desde un punto de vista de datos: Las plataformas de datos se componen de diferente tipología de datos: **Estructurados/No estructurados; Internos/Externos; Transaccionales/Analíticos**

- Orden en la arquitectura de datos** para generar confianza en los resultados
- Datawarehouse como columna vertebral**
- Enfoque piramidal de los datos

- Es necesaria una **estrategia de datos**
- Es preciso **liderazgo y cultura**
- La **organización ha de ser capaz de trabajar con datos**
- Las plataformas de datos son un **elemento clave para la transformación digital**

### Negocio

- La arquitectura tecnológica y de datos existen para satisfacer la ecuación del valor:
  - Más ventas
  - Menos costes
  - Menor circulante
  - Utilización del activo

- Los sistemas y los datos se organizan entorno a los KPIs de las áreas del negocio:
  - Financiero
  - Supply Chain y Ops
  - Marketing y Comercial
  - Recursos Humanos

- Todos los desarrollos en el ámbito de los datos y analítica han de tener un retorno claro
- Entender y trabajar con casos de uso
- Trabajar con Business Case**

## Liderazgo y Cambio cultural



# Plataformas de datos modernas | Matriz de datos

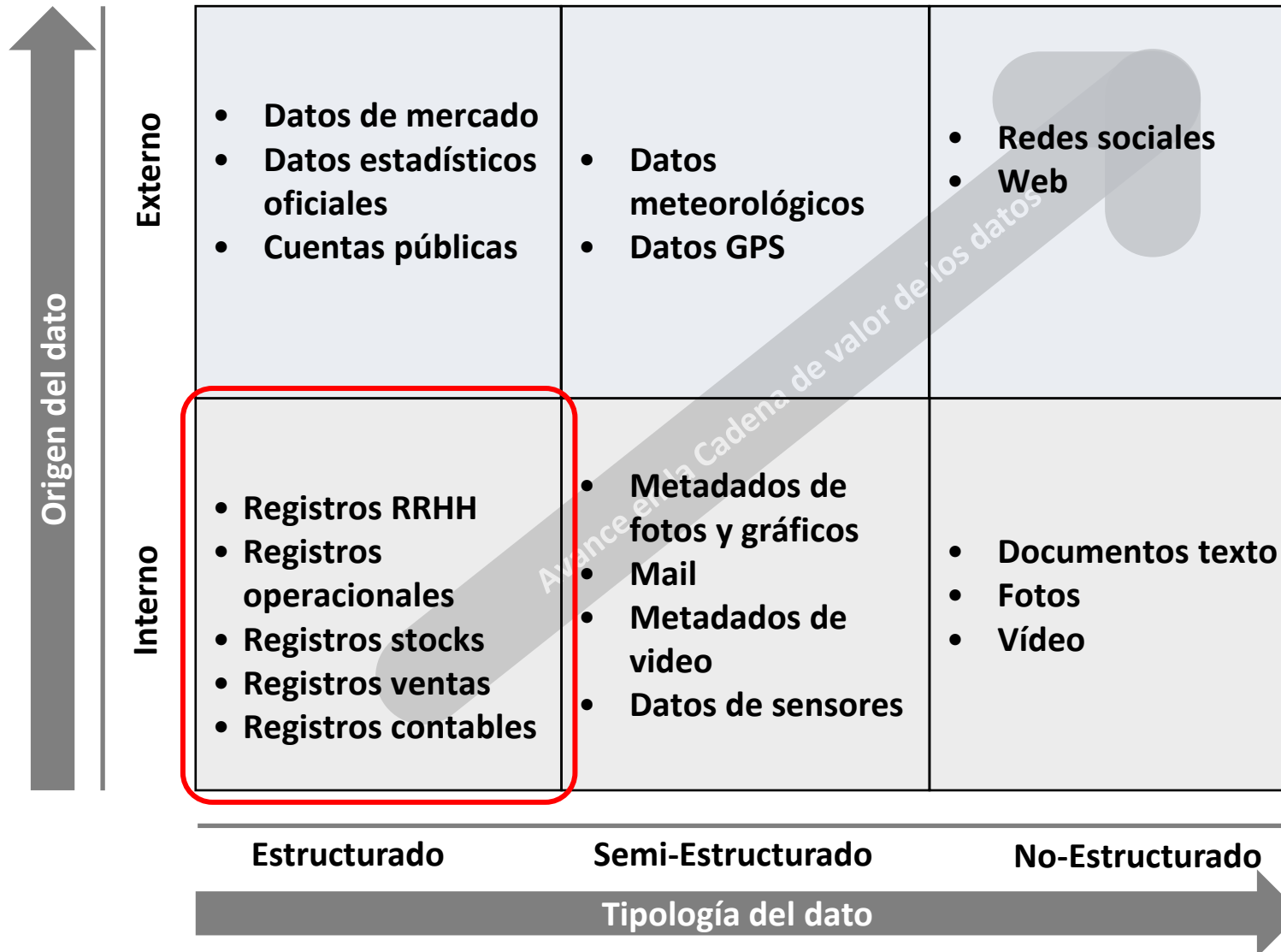
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación



Si la **arquitectura de datos** está bien definida y es **escalable**, se puede **generar conocimiento y valor** para el negocio a largo. plazo.

# Plataformas de datos modernas | Analítica Avanzada

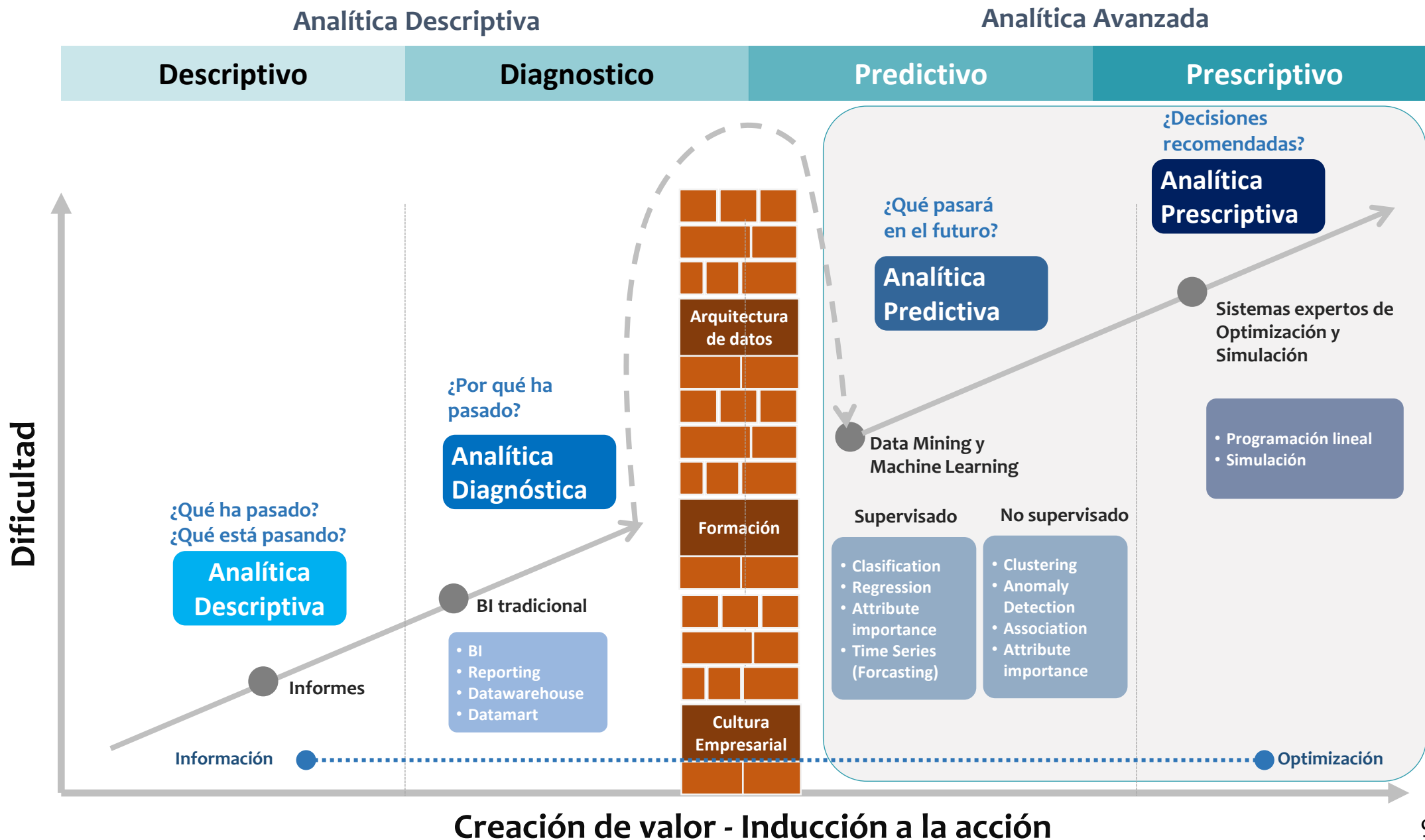
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

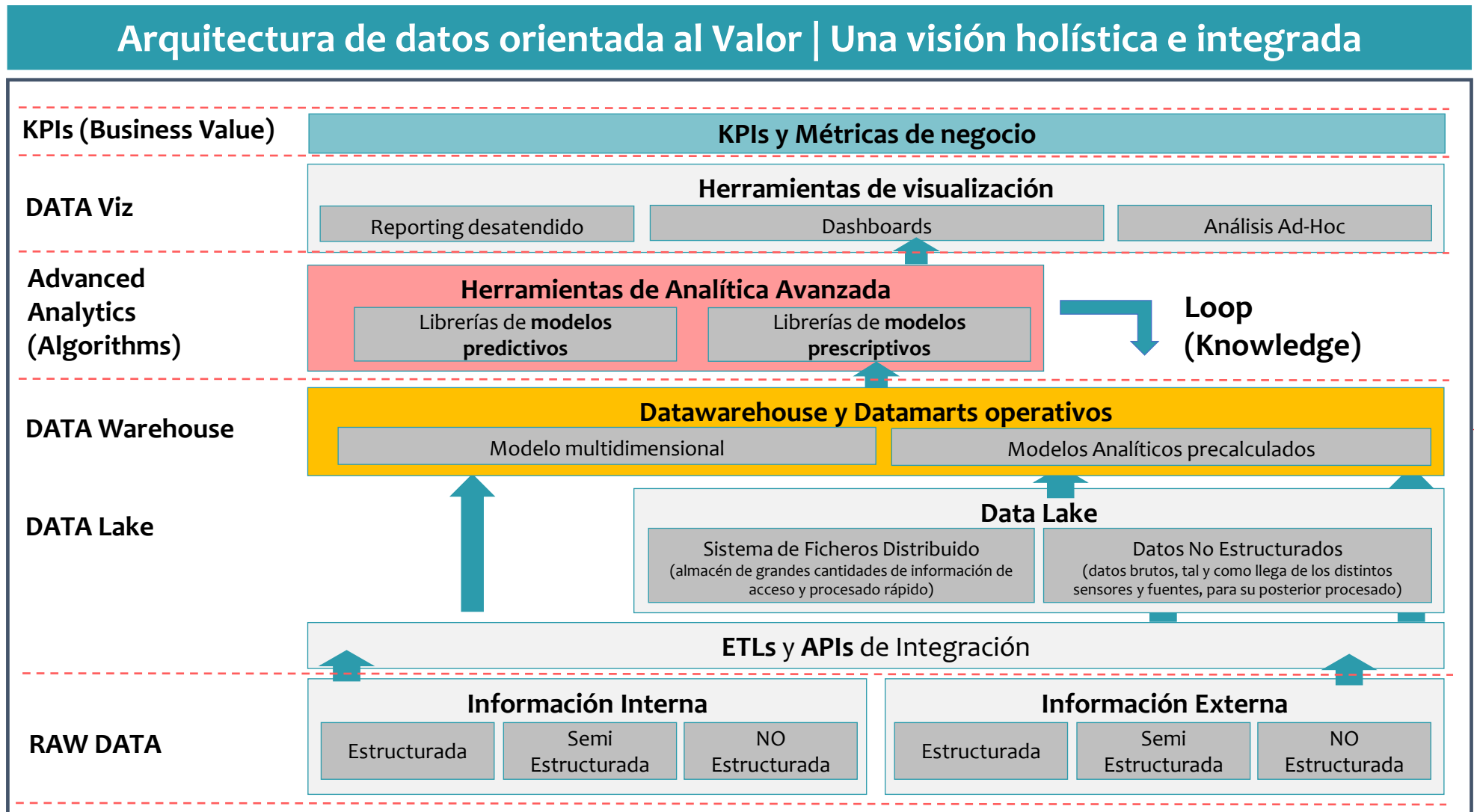
→ Proyecto de implantación



# Plataformas de datos modernas | Arquitectura

Los datos hoy en día, y cada vez más, serán una ventaja competitiva para las organizaciones.

Es precisa una visión holística e integrada de los datos.



→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

# Plataformas de datos modernas | Diferente Arquitectura

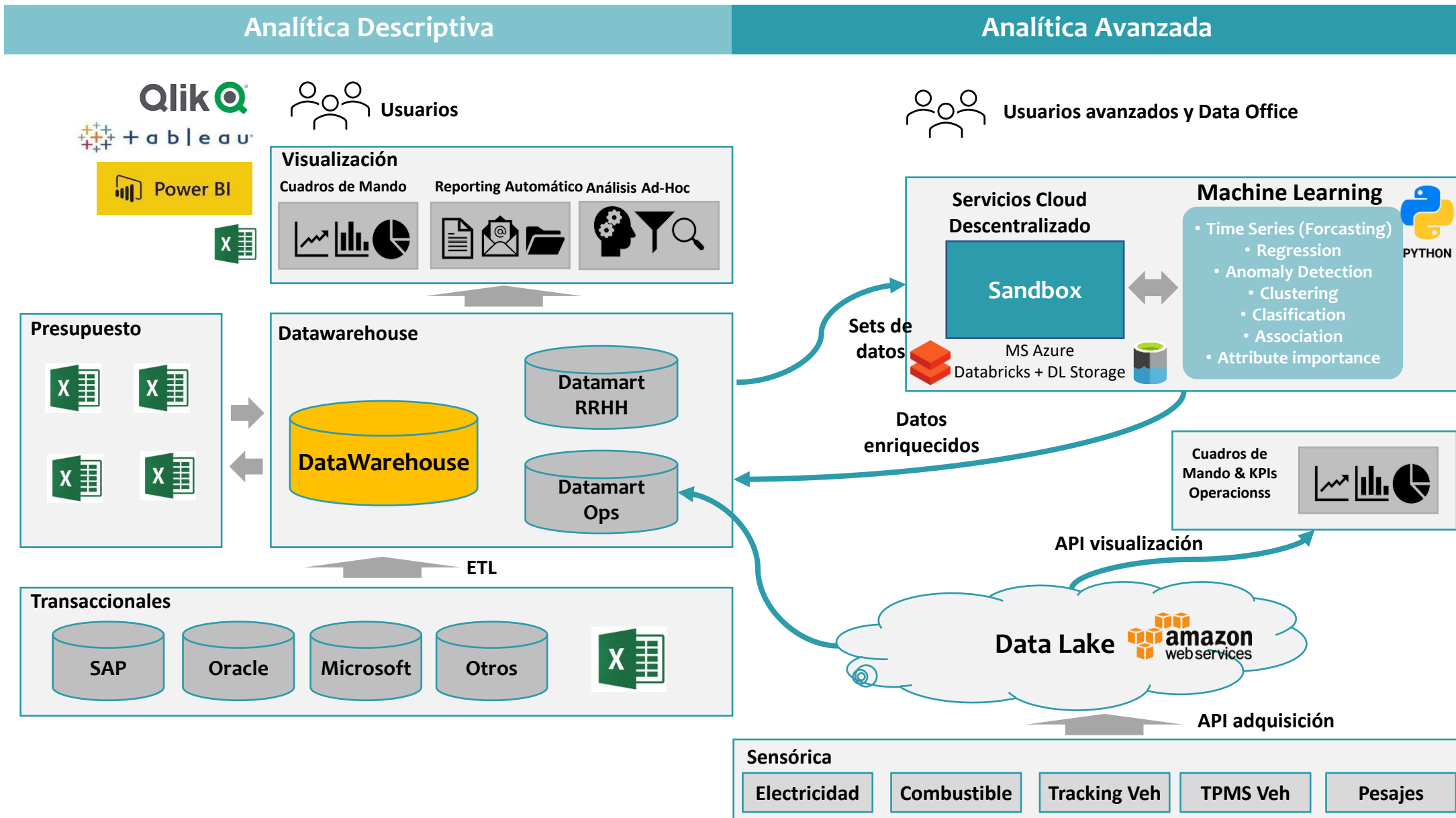
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación



# Datawarehousing

Core de la  
Plataforma

# Datawarehousing | ¿Qué es un DW?

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

**“Un Datawarehouse (DW), también conocido como almacén de datos empresarial (EDW), es un sistema utilizado para el almacenamiento de datos, estructurados de forma específica para facilitar su análisis”**

**“Un Datawarehouse es una almacén de datos que **extrae, limpia, conforma y despliega un modelo de datos multidimensional** para la consulta y el análisis”**  
**Ralph Kimball**

**El Datawarehouse permite extraer, transformar y cargar en un mismo almacén los datos de distintos orígenes (tpvs, contabilidad, crm, producción...) y establecer relaciones entre la información. Desde ese almacén los datos fluyen a Aplicaciones de Business Intelligence (BI) para crear una capa de visualización en forma de informes o dashboards”**

# Datawarehousing | ¿Qué no es un DW?

→ Plataformas de datos modernas

## → Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

- **NO ES** una copia de la base de datos “Origen” con el prefijo de nombramiento de objetos ‘DWH\_’
- **TAMPOCO** es la copia de tablas de datos (Por ej.: ‘Productos’) desde varias bases de datos y/o Plataforma de datos externas unidas todas a través de una vista
- **MUCHO MENOS** un vertedero de datos en tablas procedente de varios orígenes sin existir un diseño congruentes entre las mismas

# Plataformas de datos modernas | Diferente Arquitectura

→ Plataformas de datos modernas

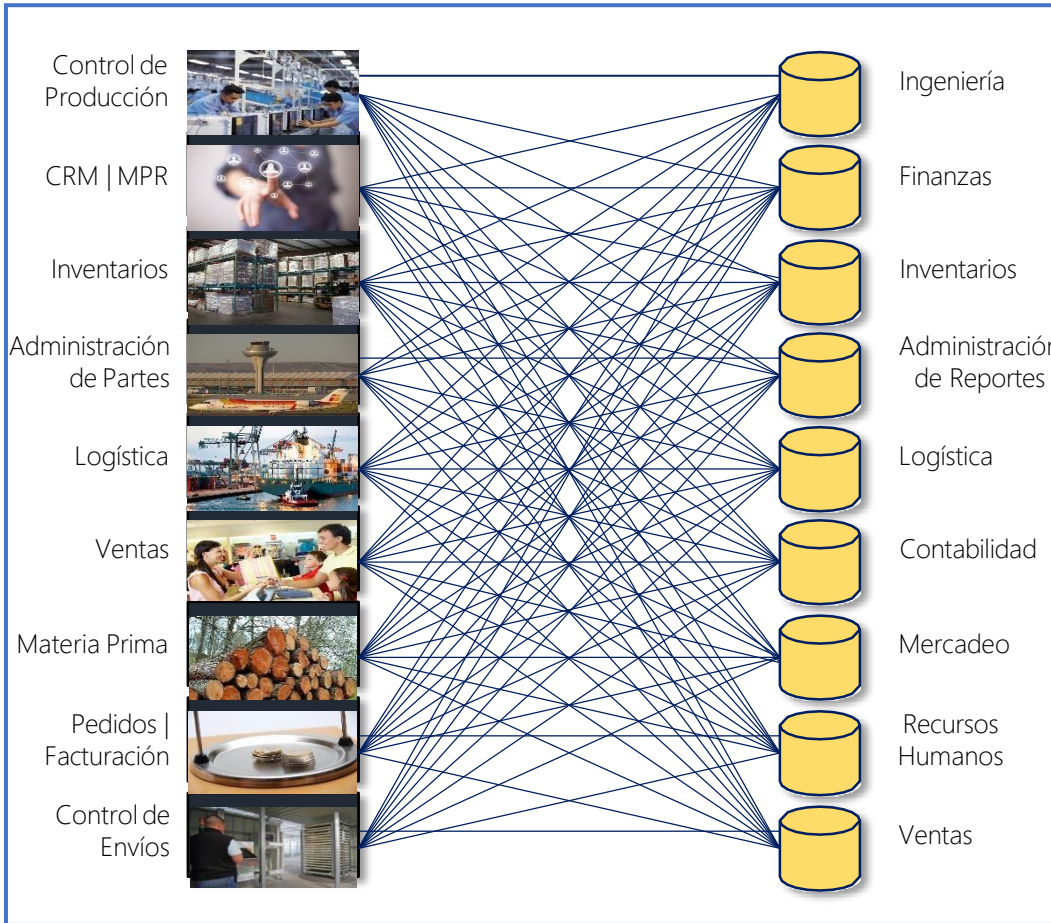
→ Datawarehousing

→ Bases de datos relacionales

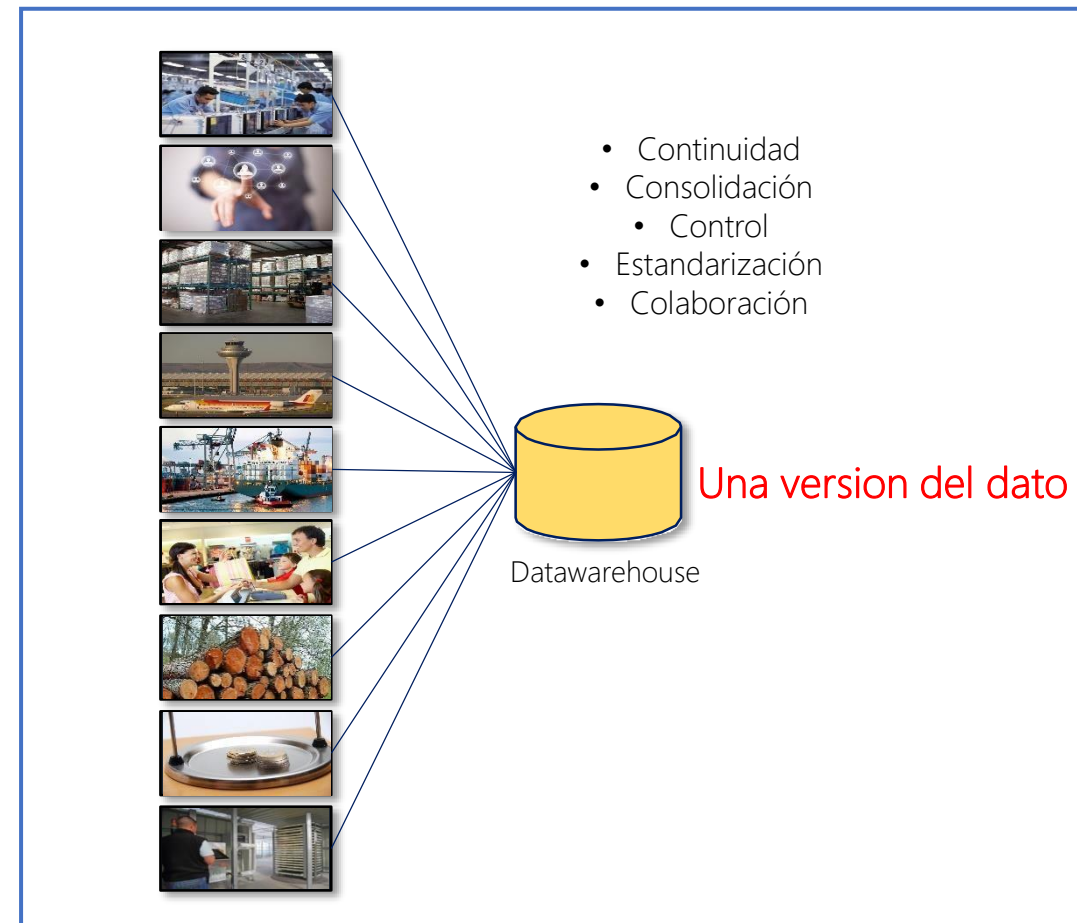
→ Modelización Multidimensional

→ Proyecto de implantación

## Múltiples consultas a los sistemas operacionales



## Enfoque Datawarehouse





# Datawarehousing | Enfoque por capas

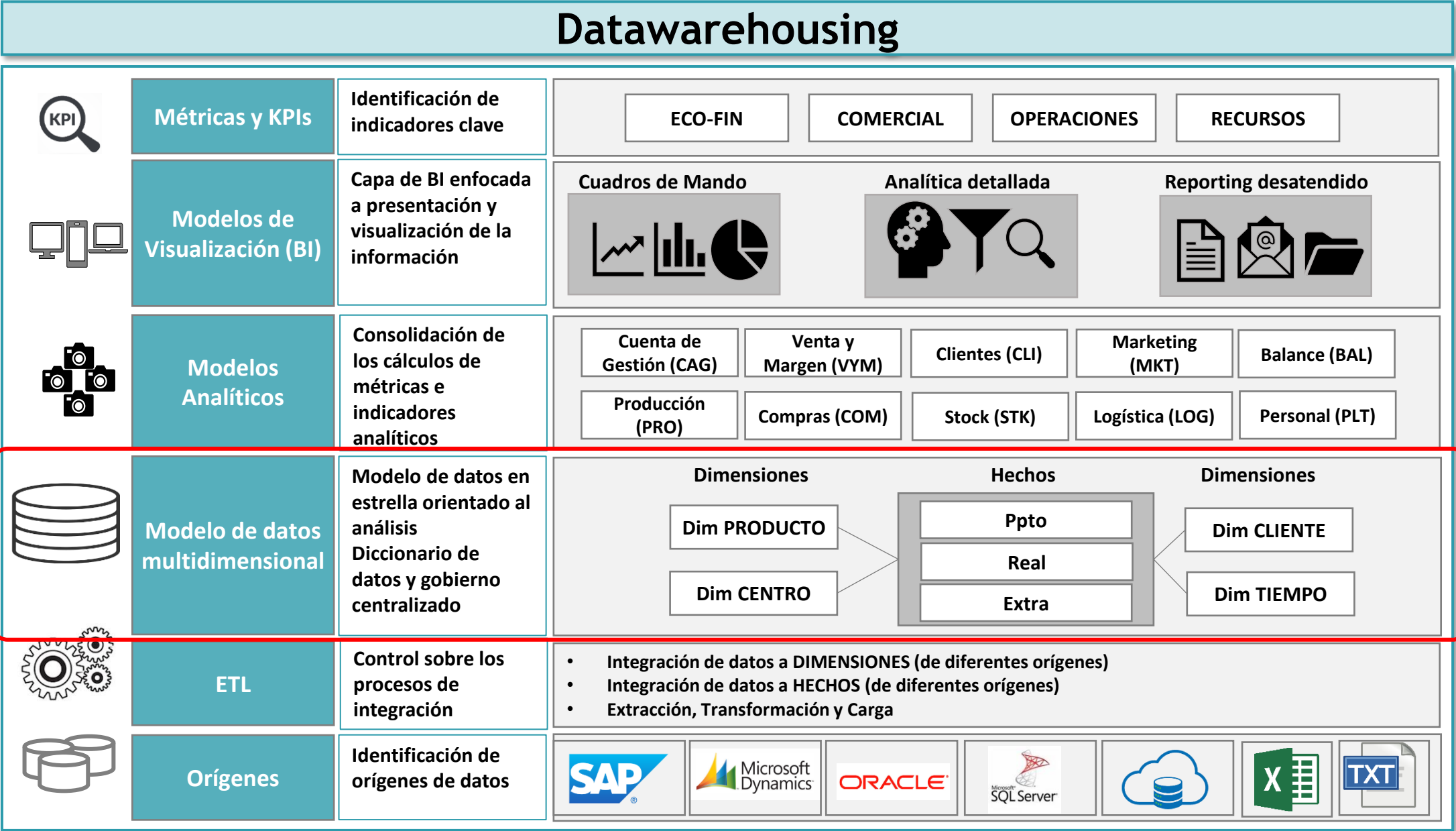
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación



# Datawarehousing | Diferente procesos

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

## EL PROCESO

Fase 1: Construir el modelo de datos



Fase 2: Definir los informes



Fase 3: Crear consultas de datos



Fase 4: Crear los informes



de la analítica descriptiva... » ...a la predictiva y prescriptiva

Qué pasó  
Cómo lo hice  
Mejoré  
Empeoré

Qué va a pasar  
Cuándo va a pasar  
Qué hacemos

Objetivo	Informes, KPIs, tendencias	Patrones y correlaciones
Proceso	Estático y comparativo	Dinámico y experimental
Fuente datos	Planificada previamente	Elaborada bajo demanda
Transformación	Previa al almacenaje	Tras el almacenaje, bajo demanda
Calidad dato	Un único dato real	Datos probabilísticos
Modelo datos	Cargado en la BBDD	Consulta sobre la BBDD
Tipo análisis	Retrospectivo, descriptivo	Predictivo, prescriptivo y preventivo

La importancia del **Data Warehouse** o almacén de datos

El **Data Warehouse** permite extraer, transformar y cargar en un mismo almacén los datos de distintos orígenes (tpvs, contabilidad, crm, producción...) y establecer relaciones entre la información. Desde ese almacén los datos fluyen a Aplicaciones de **Business Intelligence (BI)** para crear una capa de visualización en forma de informes o dashboards.

## EL PROCESO

Fase 1: Definir hipótesis a contrastar



Fase 2: Recopilar los datos



Fase 3: Construir el modelo de datos



Fase 4: Explorar los datos



Fase 5: Construir y optimizar los modelos analíticos



Fase 6: Testear la calidad del modelo



repetición

# Datawarehousing | Características

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

## Colección de datos...

**1** Orientado a temas o áreas del negocio

**2** Integrado

**3** Variante en el Tiempo

**4** No volátil

**5** Sumarizable

## ...Que facilita el proceso de toma de decisiones

### CALIDAD

Acertamos en la toma de decisiones críticas con mayor frecuencia



### VELOCIDAD

Tomamos decisiones críticas en menor tiempo



### RENDIMIENTO

Ejecutamos las decisiones críticas tal y como estaban planificadas



### ESFUERZO

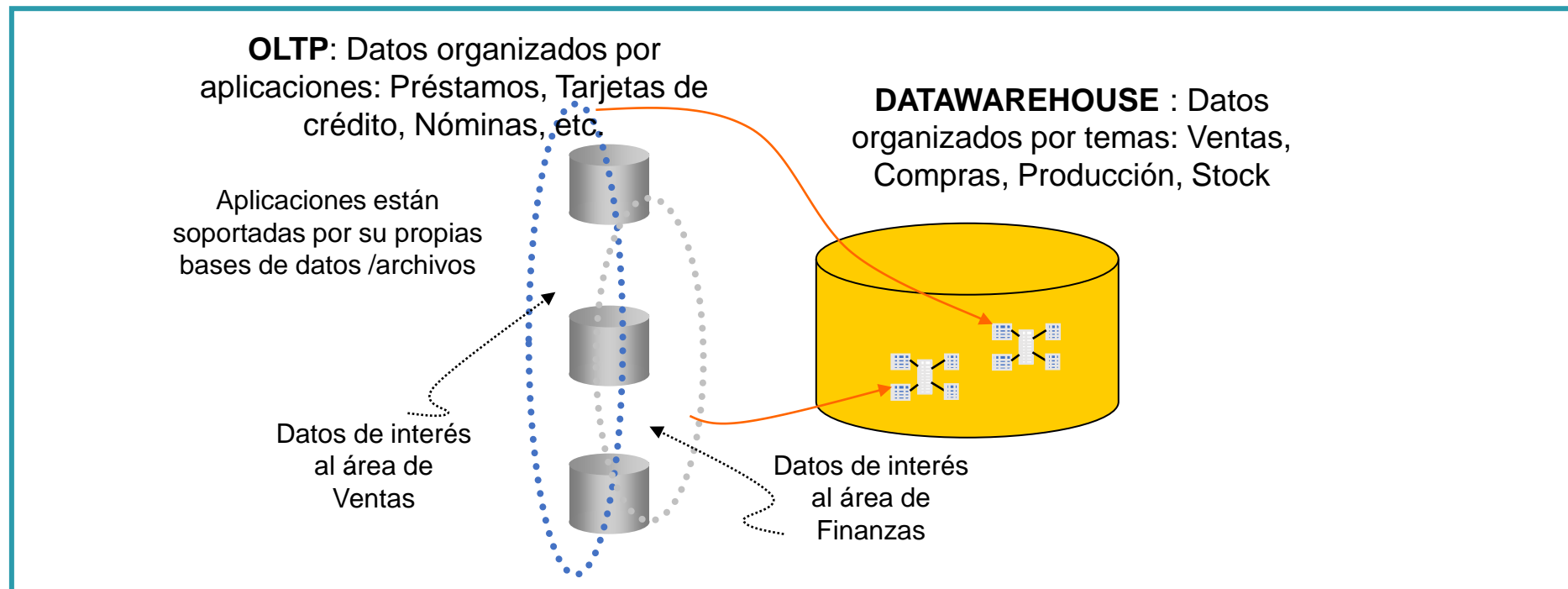
Necesitamos menos esfuerzo para la preparación y el análisis

# Datawarehousing | Características

1

## Orientado a temas o áreas del negocio

- Disponer de toda la información sobre un área
- Organización de los datos por áreas (Comercial, Producción, Stock,...)
- **La información común a varios temas no debe duplicarse**



→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

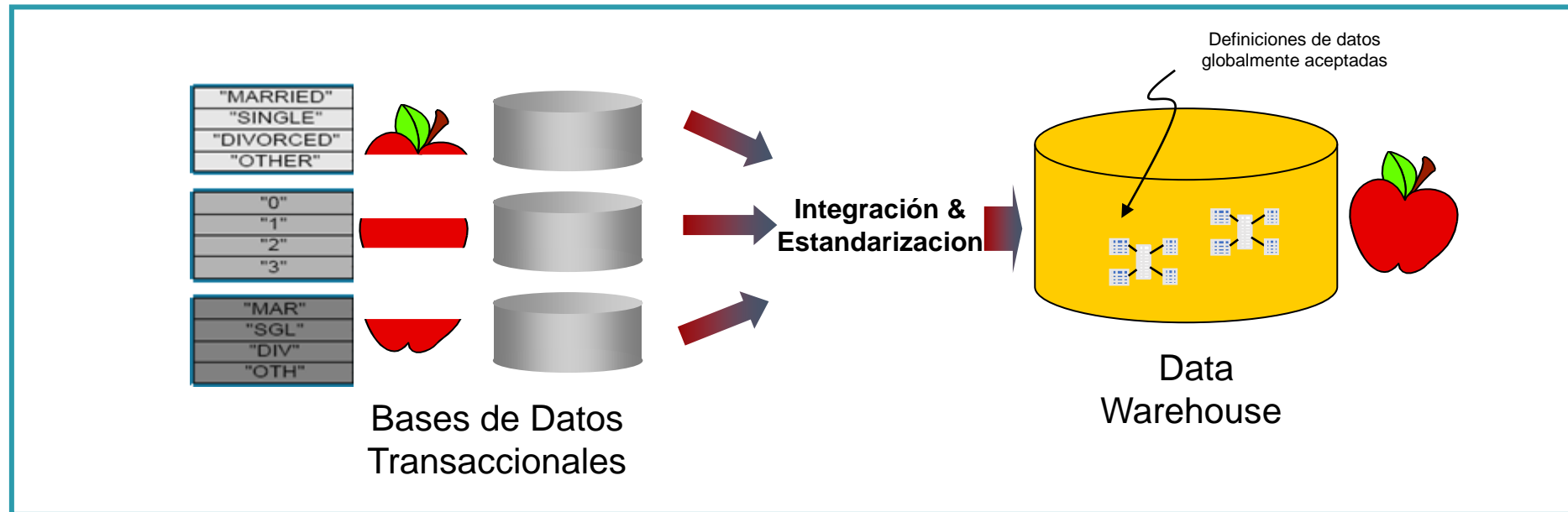
→ Proyecto de implantación

# Datawarehousing | Características

2

## Integrado

- La información se almacena en el DW de acuerdo a convenciones globales (aún cuando la fuente que genera los datos se almacena de forma diferente).
- **La semántica de medidas, dimensiones, etc. es global y homogénea.**
- Los datos deben formatearse y unificarse para llegar a un estado consistente.



→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

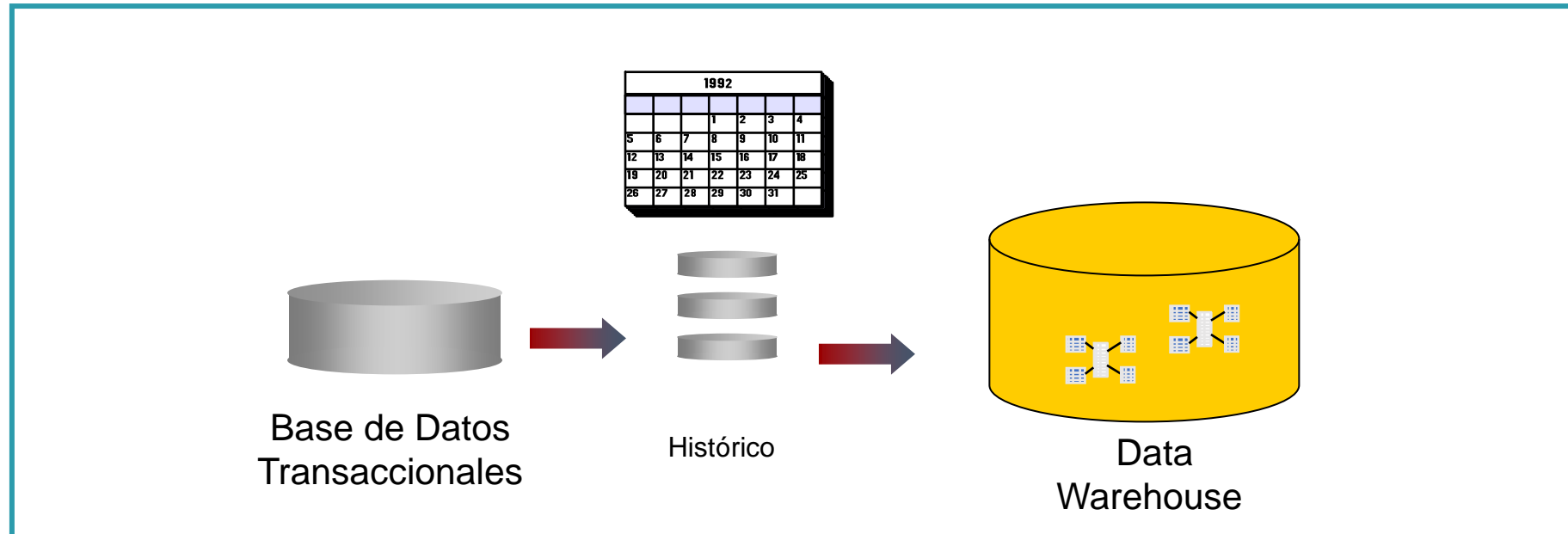
→ Proyecto de implantación

# Datawarehousing | Características

3

## Variante en el Tiempo

- La información del Datawarehouse se utiliza a largo plazo, normalmente por un período de cinco a diez años.
- Los datos **se almacenan como un conjunto de fotos instantáneas donde cada una representa a un período de tiempo.**
- Los datos no se actualizan nunca. Representan un valor en un momento concreto
- Los datos se referencian temporalmente.



→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

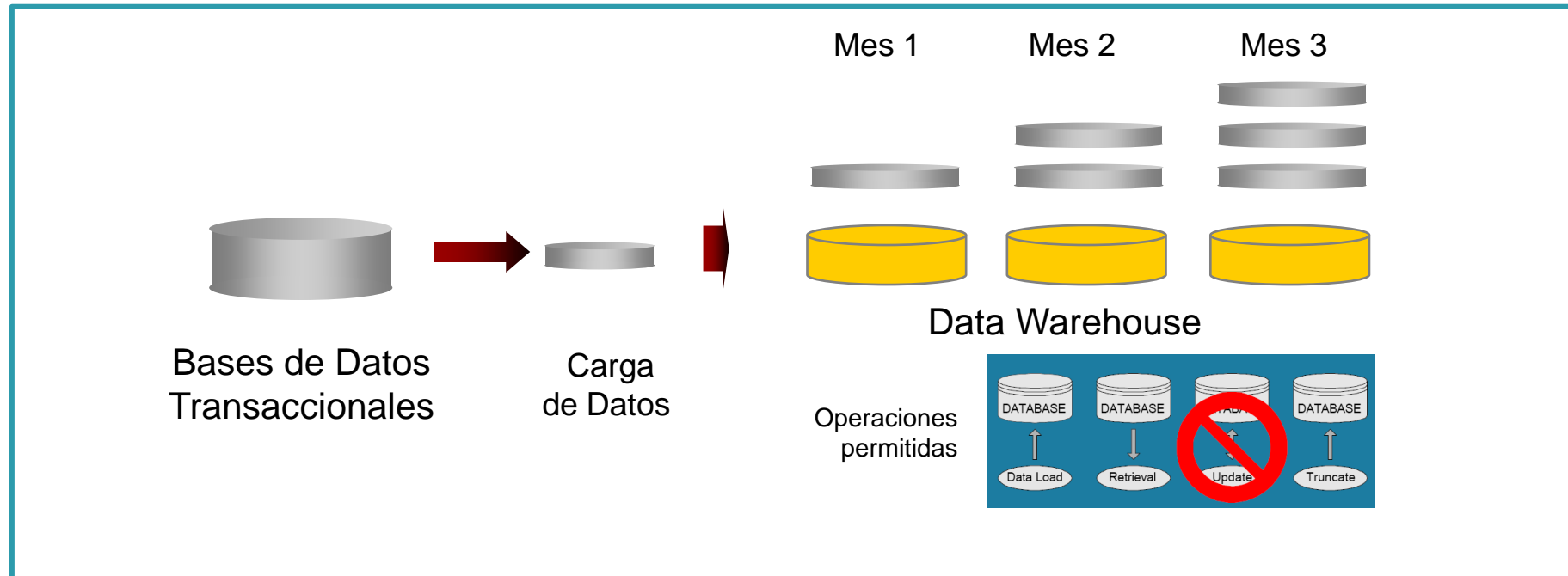
→ Proyecto de implantación

# Datawarehousing | Características

4

## No volátil

- **La información en el Datawarehouse es sólo de lectura.**
- Los datos se cargan dentro del Datawarehouse y se acceden ahí mismo.
- Una vez que se toma una fotografía instantánea de los datos, estos no cambian dentro del Datawarehouse.



# Datawarehousing | Diferentes enfoque de BBDD

→ Plataformas de datos modernas

→ Datawarehousing

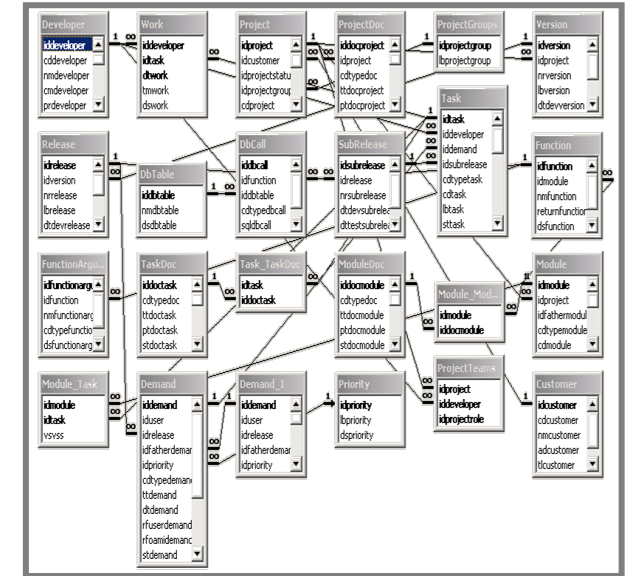
→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

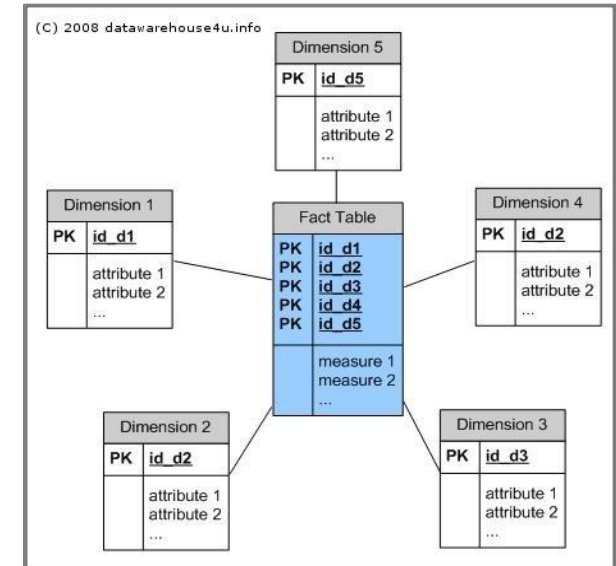
Bases de datos transaccionales

- Datos organizados por aplicaciones
- Datos muy normalizados (muchas tablas)
- Lectura / Escritura
- Crecimiento constante
- Información actual



Bases de datos analíticas

- Datos organizados por temas
- Datos desnormalizados (y pocas tablas)
- Frecuentemente lectura
- Actualización masiva periódica
- Información histórica





# Datawarehousing | Operacional vs Datawarehouse

→ Plataformas de datos modernas

## → Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

Comparación	Operacional	Datawarehouse
<b>Datos</b>	Transaccionales	Del negocio
<b>Uso de los datos</b>	Procesamiento repetitivo	Procesamiento analítico
<b>Orientación del diseño</b>	A la aplicación	Al tema (Comercial, Stock,..)
<b>Estructura de los datos</b>	Muchas tablas normalizadas	Pocas y desnormalizadas
<b>Datos en el tiempo</b>	Actuales	Actuales+Históricos+Ppto
<b>Detalle de los datos</b>	Altamente detallados	Detallados y Agregados
<b>Cambios en los datos</b>	Continuos	Refrescos periódicos
<b>Cantidad de usuarios</b>	Más que un DW	Menos que un transaccional
<b>Tamaño de la BBDD</b>	100 MB -CB	100 GB-TB
<b>Registros accedidos en una operación</b>	Decenas	Millones

# Datawarehousing | Datawarehouse vs Datamart

→ Plataformas de datos modernas

## → Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

Característica	DATAWAREHOUSE	DATAMART
Ámbito de los datos	Corporativo	Departamental
Número procesos	Múltiples	Simple
Dificultad	Alta	Media
Tiempo construcción	Largo	Medio
Memoria	Alta	Limitada

# Datawarehousing | Arquitectura funcional

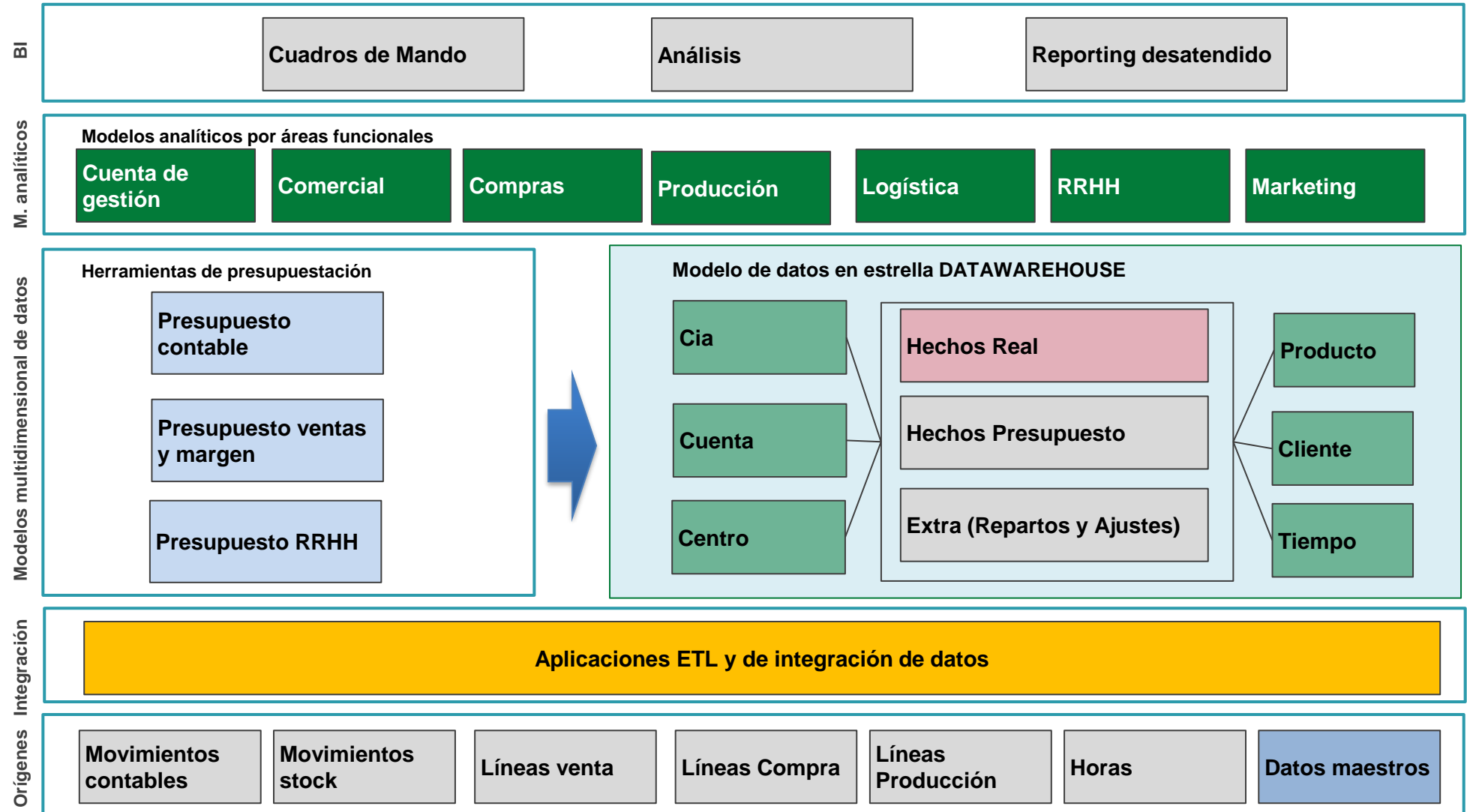
→ Plataformas de datos modernas

## → Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación



# Datawarehousing | Arquitectura tecnológica

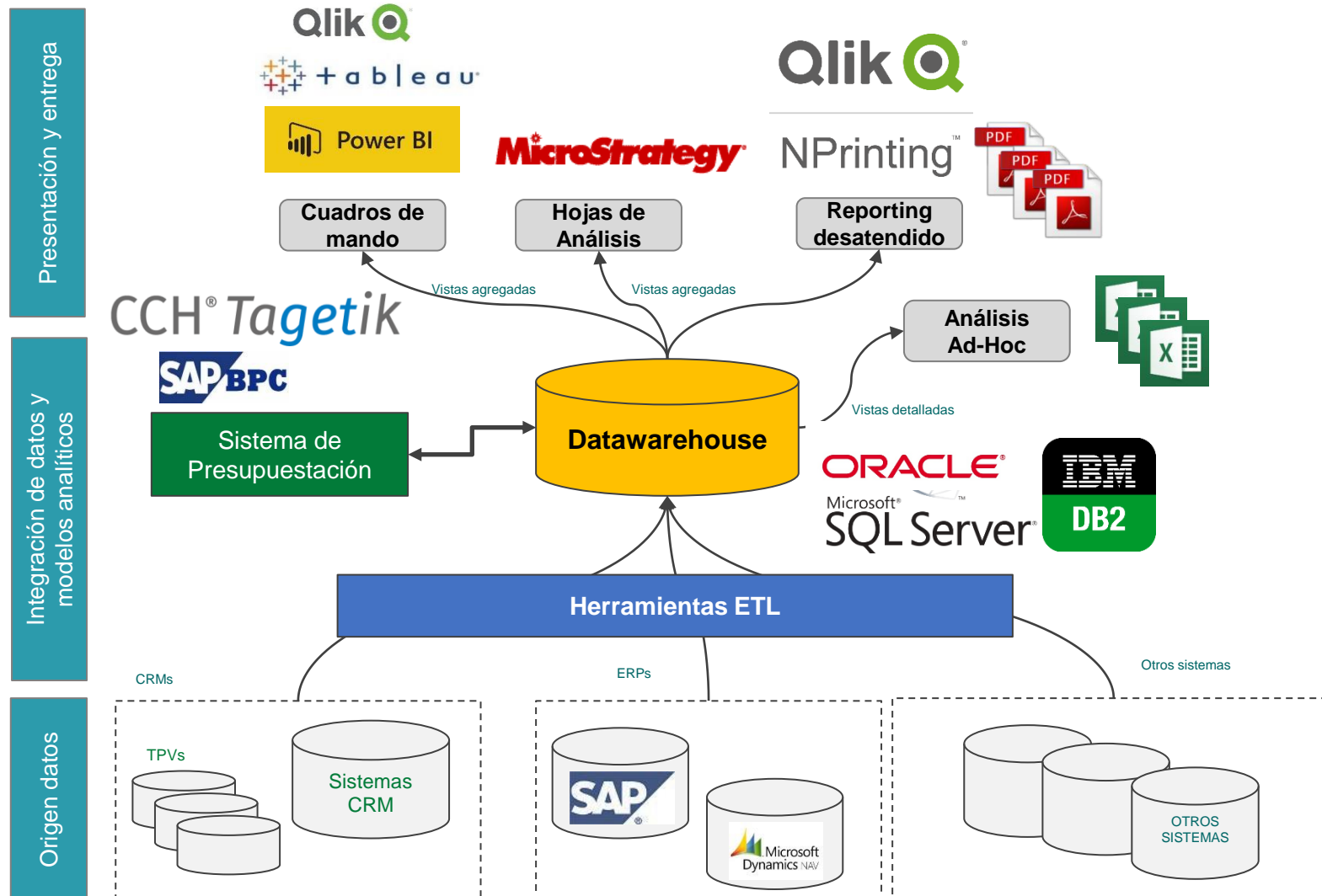
→ Plataformas de datos modernas

## → Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

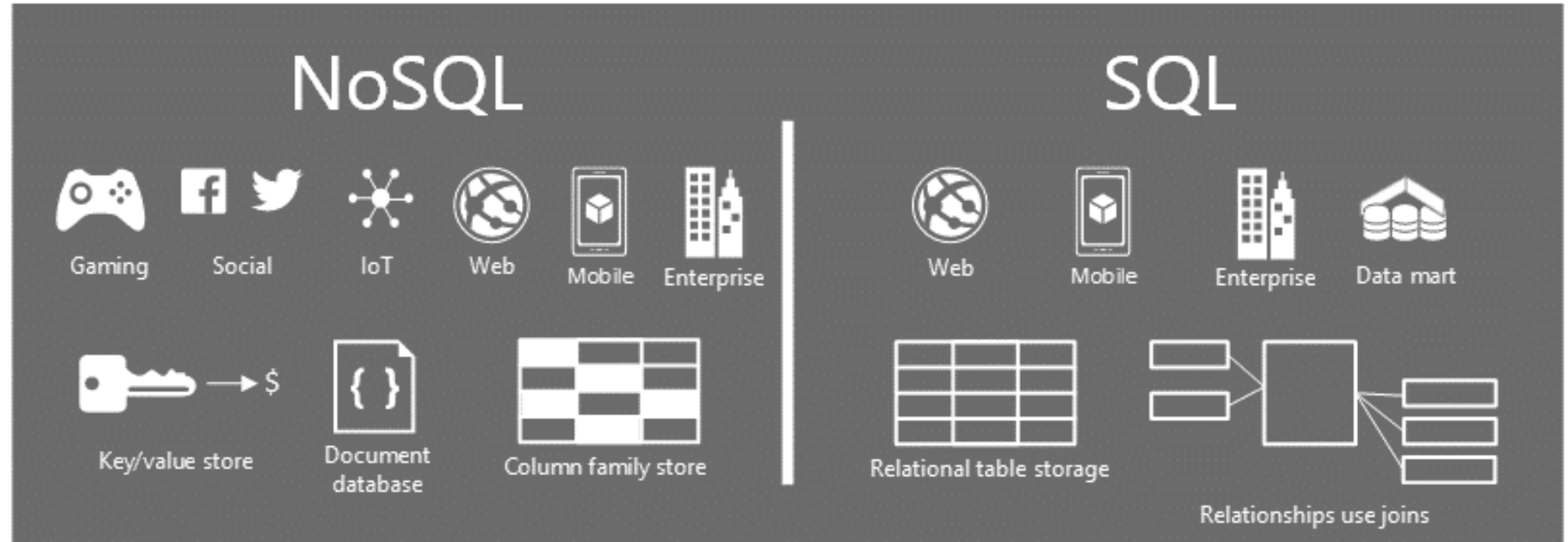


# Bases de datos relacionales

Fundamentos

# Bases de datos relacionales | Contexto

El mundo de las bases de datos se divide fundamentalmente en 2 tipologías:



Procesamiento distribuido  
Escalabilidad  
Datos no estructurados

Uso extendido  
Soporte  
Atomicidad de operaciones  
Datos estructurados

# Bases de datos relacionales | Elementos principales

## Tablas

- **Una tabla es una estructura lógica que sirve para almacenar los datos de un mismo tipo (desde el punto de vista conceptual).** Desde el punto de vista conceptual esto significa que cada entidad se almacena en estructuras separadas.
  - Por ejemplo: la entidad factura se almacena en estructuras diseñadas para ese tipo de entidad: la tabla FACTURA y la tabla FACTURA\_COMPRA etc. Así, cada entidad, tendrá una estructura (tabla) pensada y diseñada para ese tipo de entidad.
- Cada entidad almacenada dentro de la tabla recibe el nombre de registro o fila. Así si la tabla FACTURA almacena 1.000 facturas, se dice que la tabla FACTURA contiene 1.000 registros o filas.
- Una tabla se compone de campos o columnas, que son conjuntos de datos del mismo tipo (desde el punto de vista físico).
  - Ahora cuando decimos “del mismo tipo” queremos decir que los datos de una columna son de todos del mismo tipo: numéricos, alfanuméricos, fechas...

Nº factura	Descripción	Cliente	Importe	%Descuento	Importe final
001	Tornillos sin rosca	Pepe	1.000	10	900
002	Tuercas sin agujero	Juancito	5.500	0	5.500
003	Tuercas de segunda mano	Toñete	500	1	495

Filas o registros

Columnas o campos

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

# Bases de datos relacionales | Elementos principales

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

- **Claves primarias**
  - Una clave primaria es una columna o un conjunto de columnas en una tabla cuyos valores identifican de forma exclusiva una fila de la tabla
  - Cada tabla debe poseer una **clave primaria**, esto es, un **identificador único de cada registro compuesto por una o más columnas**.
- **Clave foránea**
  - Una clave foránea es una columna o un conjunto de columnas en una tabla cuyos valores corresponden a los valores de la clave primaria de otra tabla.
  - Para establecer una relación entre dos tablas es necesario incluir, en forma de columna, en una de ellas la clave primaria de la otra. A esta columna se le llama clave foránea.
- **Relaciones entre tablas**
  - Relación 1 a 1
  - Relación 1 a n
  - Relación n a n
- **Diseño de bases de datos**
  - **Diseño conceptual** (estructuras de datos a alto nivel)
  - **Diseño lógico** (definir las tablas que existirán y las relaciones entre ellas)
  - **Diseño físico** (la forma de almacenamiento)



# Bases de datos relacionales | Ejercicio 1

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

Northwind
Tables
dbo.Categories
dbo.CustomerDemographics
dbo.Customers
dbo.Employees
dbo.Order Details
dbo.Orders
dbo.Products
dbo.Region
dbo.Shippers
dbo.Suppliers

Categories

Column Name	Data Type
CategoryID	int
CategoryName	nvarchar(15)
Description	ntext
Picture	image

Customers

Column Name	Data Type
CustomerID	nchar(5)
CompanyName	nvarchar(40)
ContactName	nvarchar(30)
ContactTitle	nvarchar(30)
Address	nvarchar(60)
City	nvarchar(15)
Region	nvarchar(15)
PostalCode	nvarchar(10)
Country	nvarchar(15)
Phone	nvarchar(24)
Fax	nvarchar(24)

Employees

Column Name	Data Type
EmployeeID	int
LastName	nvarchar(20)
FirstName	nvarchar(10)
Title	nvarchar(30)
TitleOfCourtesy	nvarchar(25)
BirthDate	datetime
HireDate	datetime
Address	nvarchar(60)
City	nvarchar(15)
Region	nvarchar(15)
PostalCode	nvarchar(10)
Country	nvarchar(15)
HomePhone	nvarchar(24)
Extension	nvarchar(4)
Photo	image
Notes	ntext
ReportsTo	int
PhotoPath	nvarchar(255)

Order Details

Column Name	Data Type
OrderID	int
ProductID	int
UnitPrice	money
Quantity	smallint
Discount	real

Orders

Column Name	Data Type
OrderID	int
CustomerID	nchar(5)
EmployeeID	int
OrderDate	datetime
RequiredDate	datetime
ShippedDate	datetime
ShipVia	int
Freight	money
ShipName	nvarchar(40)
ShipAddress	nvarchar(60)
ShipCity	nvarchar(15)
ShipRegion	nvarchar(15)
ShipPostalCode	nvarchar(10)
ShipCountry	nvarchar(15)

Products

Column Name	Data Type
ProductID	int
ProductName	nvarchar(40)
SupplierID	int
CategoryID	int
QuantityPerUnit	nvarchar(20)
UnitPrice	money
UnitsInStock	smallint
UnitsOnOrder	smallint
ReorderLevel	smallint
Discontinued	bit

Region

Column Name	Data Type
RegionID	int
RegionDescription	nchar(50)

Shippers

Column Name	Data Type
ShipperID	int
CompanyName	nvarchar(40)
Phone	nvarchar(24)

Suppliers

Column Name	Data Type
SupplierID	int
CompanyName	nvarchar(40)
ContactName	nvarchar(30)
ContactTitle	nvarchar(30)
Address	nvarchar(60)
City	nvarchar(15)
Region	nvarchar(15)
PostalCode	nvarchar(10)
Country	nvarchar(15)
Phone	nvarchar(24)
Fax	nvarchar(24)
HomePage	ntext

**Ejercicio:** Identificar las claves primarias de cada tabla y dibujar el Diagrama Entidad Relación de la Base de datos

# Bases de datos relacionales | Ejercicio 1

→ Plataformas de datos modernas

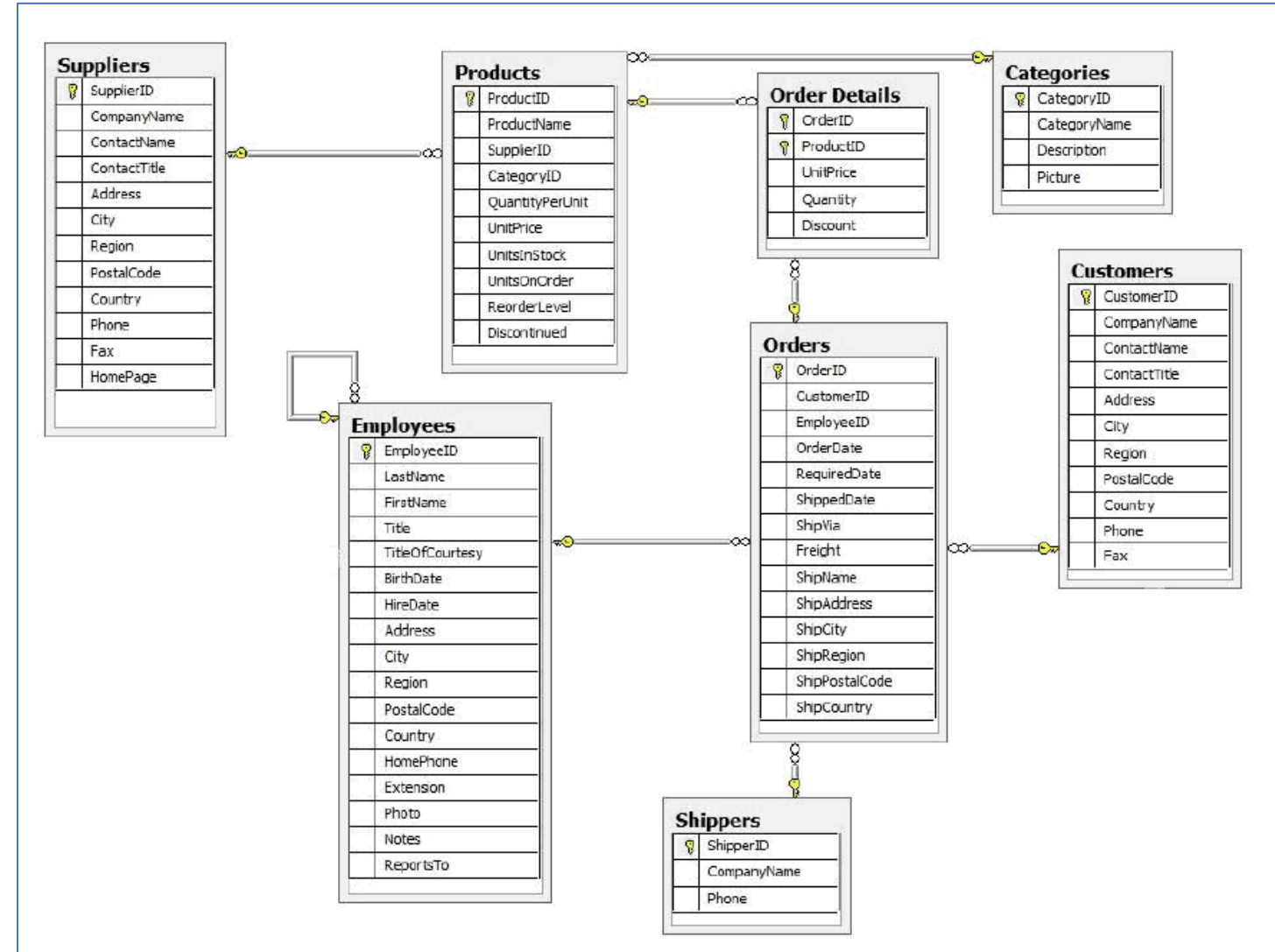
→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

Diagrama Entidad Relación (E-R) de la base de datos Northwind



# Bases de datos relacionales | Ejercicio 2

→ Plataformas de datos modernas

→ Datawarehousing

→ **Bases de datos relacionales**

→ Modelización Multidimensional

→ Proyecto de implantación

## Consultas a la base de datos y generación de informes (basadas en SQL)

1. Mostrar todas las columnas y registros de la tabla “Region”
2. Listar todos los empleados en formato Apellido, Nombre, ordenando ascendente por apellido
3. Listar todos los pedidos ordenados en descendente por coste transporte (mostrar también el id de pedido, la fecha del pedido, fecha de entrega y el cliente)
4. Crear un informe que muestre el título, nombre y apellido de todos los empleados que sean representantes de ventas.
5. Crear un informe que muestre el nombre y apellido de todos los empleados que tienen una región especificada.
6. Crear un informe que muestre el nombre y apellido de todos los representantes de ventas que son de Seattle o Redmond.
7. Calcular el número total de unidades pedidas del producto 3.
8. Calcular el número de empleados que hay por cada ciudad.
9. Crear un informe que liste todos los pedidos de los empleados ordenados por Empleado y Fecha
10. **Crear un informe (set de datos) que liste toda la siguiente información:**
  - Año de la venta
  - Mes de la venta
  - Empleado venta (nombre y apellidos)
  - Posición (título del empleado)
  - Producto (código)
  - Producto (nombre)
  - Categoría del producto (nombre)
  - Nombre proveedor del producto
  - País del proveedor
  - Cliente (nombre de la empresa)
  - País del cliente
  - Coste transporte
  - Cantidad de unidades
  - Importe venta bruto (sin descuento)
  - Importe venta neto (con descuento)

# Bases de datos relacionales | Ejercicio 2

Generar un análisis dinámico (Excel) a partir del set de datos anterior que muestre:

- Ventas netas por categoría y mes
- Ventas netas por país y mes
- Top de productos por categoría (por año)
- Ranking de clientes (por año)
- Ranking de empleados (por ventas) por año
- Ranking de proveedores por categoría

→ Plataformas de datos modernas

→ Datawarehousing

→ **Bases de datos relacionales**

→ Modelización Multidimensional

→ Proyecto de implantación

# Modelización Multidimensional

## Fundamentos

# Modelización Multidimensional | Contexto

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

Si en una empresa comercializadora nos pasan un Informe de que las ventas son:

157.234 €

Seguramente se nos interesará analizar este importe de ventas (**Medidas**) por algún criterio analítico (**Dimensiones**):

Dimensión	Pensamiento dimensional	Ejemplo
Producto	¿Qué se vendió?	Productos de una determinada Línea (Aceite, Detergente, Lácteos)
Cliente	¿A quién se vendió?	Clientes de un Determinado Segmento (Regulares, Principales, Esporádicos)
Comercial	¿Quién lo vendió?	Ranking de ventas por comercial Comercial A, Comercial B,...
Centro	¿Dónde se vendió?	Comparativo de ventas entre centros Centros Norte, Centros Sur
Comercial	¿Quién lo vendió?	Ranking de ventas por comercial Comercial A, Comercial B,...
Canal	¿A través de qué canal?	Distribución de ventas por canal (industrial, distribución, on-line,..)

# Modelización Multidimensional

**Los hechos contienen los valores de las métricas, medidas o indicadores de negocio.**

Son tablas que tienen **dos tipos de campos, las que apuntan a las dimensiones (claves foráneas) o las que contienen las medidas (valores, magnitudes)**

**Hechos**

- Hechos contables
- Hechos de ventas
- Hechos de compras
- Hechos de producción
- Hechos de stock
- Hechos de finanzas
- Hechos de RRHH
- Hechos de TI

**Métricas y Medidas**

Constituyen el qué analizar, son numéricas y sumabilizables a diferente nivel de detalle.

**Las medidas se agrupan en métricas**

**Dimensiones**

**Son perspectivas de análisis y determinan el como analizar a las medidas.**

Por ejemplo, la dimensión cliente (quién), producto (qué), tiempo (cuándo), ubicación (dónde), canal (cómo), entre otras.

La granularidad de la tabla de hechos es la clave única (conjunto de dimensiones)

**Atributos**

**Son propiedades de los elementos o niveles de la dimensión.**

Algunos atributos son para mostrar (descripciones) otros son como tamaño, color,...

**Jerarquías**

Una jerarquía implica una organización de niveles dentro de una dimensión.

Una dimensión debe contener al menos una jerarquía. Esta puede contener a su vez uno o más niveles. Por ejemplo, la dimensión cliente puede tener una jerarquía que organiza los clientes por su ubicación geográfica y otra que los organiza por capacidad de compra

**Niveles**

Si una dimensión contiene una jerarquía que se ha estructurado tipo árbol, se denominará nivel a cada peldaño de la estructura.

→ Plataformas de datos modernas

→ Datawarehousing

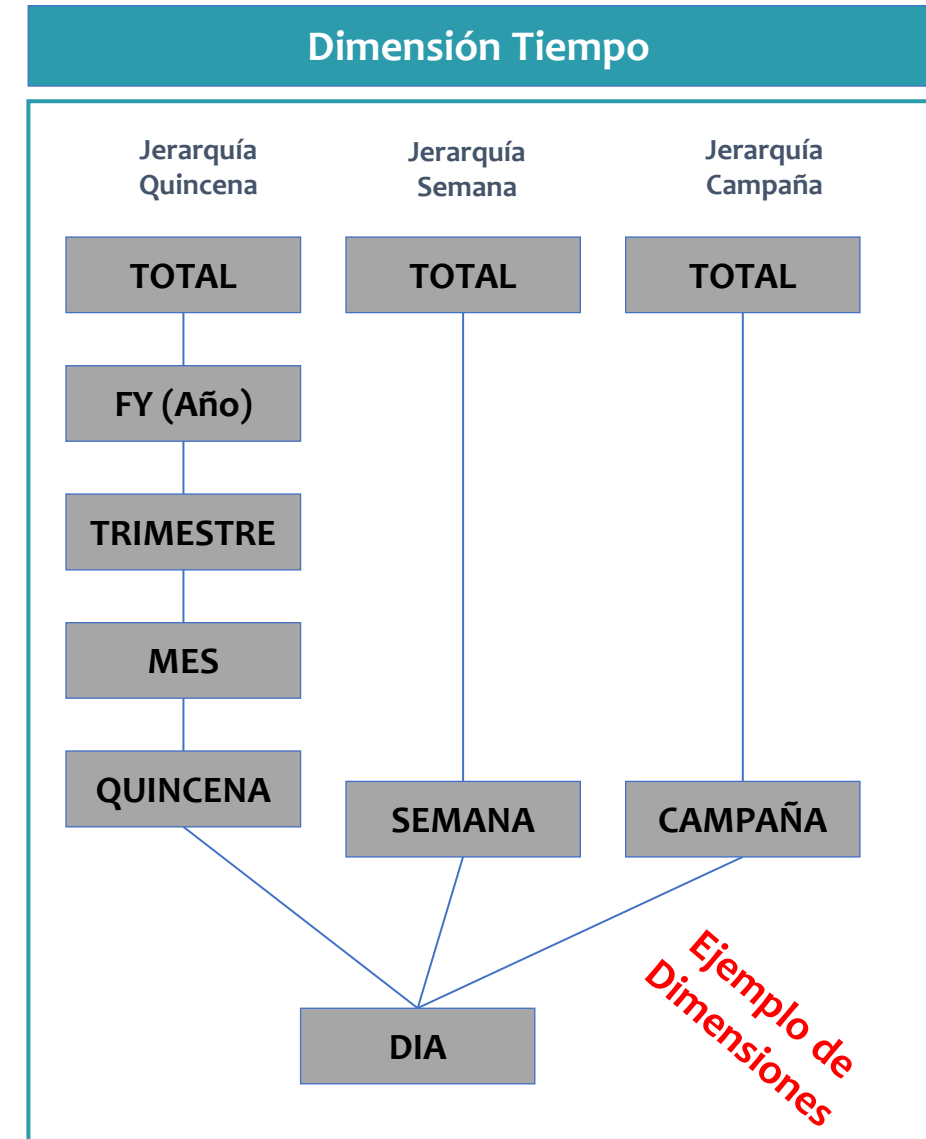
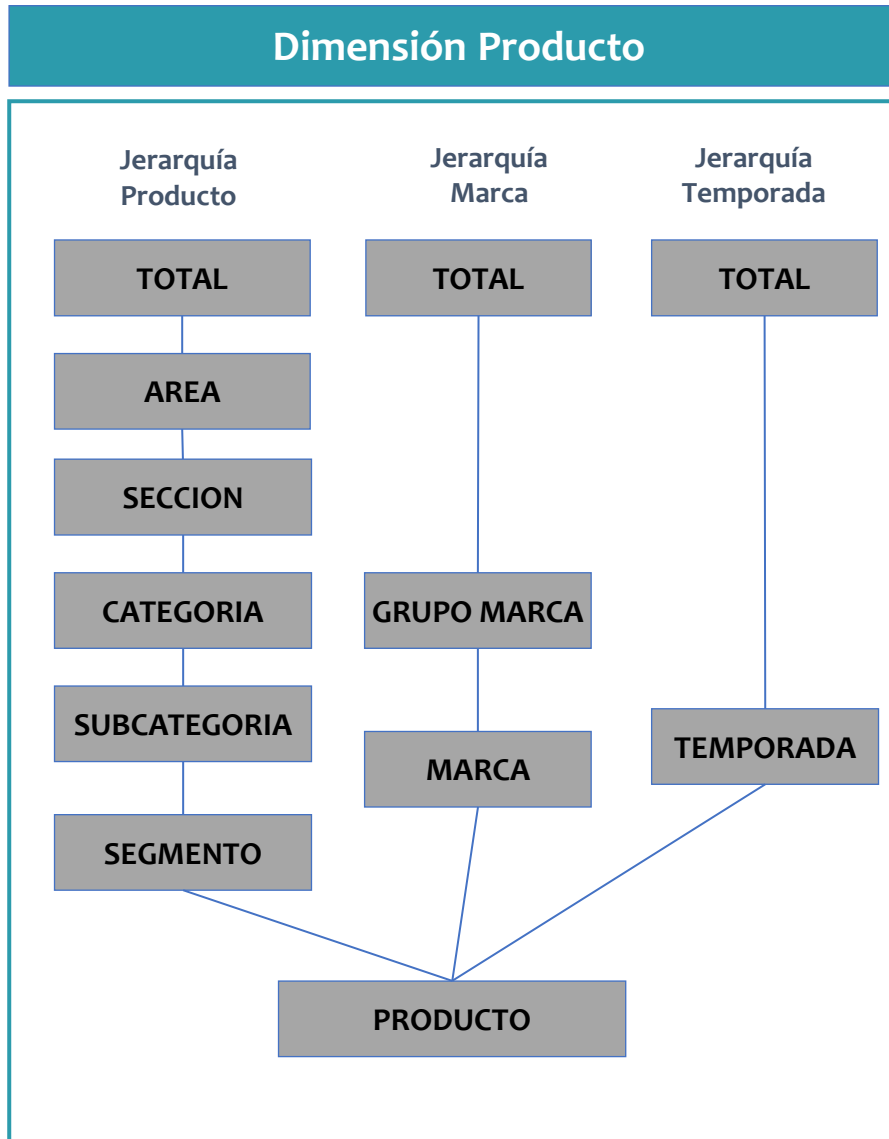
→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

# Modelización Multidimensional

## Ejemplos de dimensiones:



→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación



# Modelización Multidimensional

## Ejemplos de métricas y medidas:

Métrica SRV   Nivel de servicio	
<b>Tiempo de entrega</b>	Tiempo medio de la entrega del producto/servicio
<b>% Pedidos correctos en tiempo (OTD)</b>	% de pedidos sobre el total de pedidos entregados en tiempo
<b>% Pedidos correctos en cantidad</b>	% de pedidos sobre el total de pedidos con las cantidades correctas
<b>% Pedidos facturados correctamente</b>	% de pedidos sobre el total de pedidos facturados de forma correcta
<b>% Pedidos servidos perfectos (POF)</b>	% de pedidos perfectos sobre el total de pedidos entregados (multiplicación)
<b># Reclamaciones</b>	Número total de reclamaciones realizadas
<b># Reclamaciones por servicio o producto</b>	Número de reclamaciones sobre servicio o producto
<b>€ en reclamaciones</b>	Importe en € de las reclamaciones recibidas sobre producto o servicio
<b>€ reclamados / € vendidos</b>	% de importe reclamado entre el importe total vendido
<b>€ reclamados / Unidades vendidas</b>	€ reclamados por unidad vendida
<b># Devoluciones</b>	Número total de devoluciones realizadas
<b>€ de las devoluciones recibidas</b>	Importe en € de las devoluciones recibidas
<b>% devoluciones</b>	% de las devoluciones en € entre el total de ventas en €
<b># incidencias</b>	Incidencias de cliente registradas
<b>Tiempo de resolución de incidencias</b>	Tiempo medio para la resolución de incidencias desde apertura a cierre

*Ejemplo de Métricas y medidas*

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

# Modelización Multidimensional

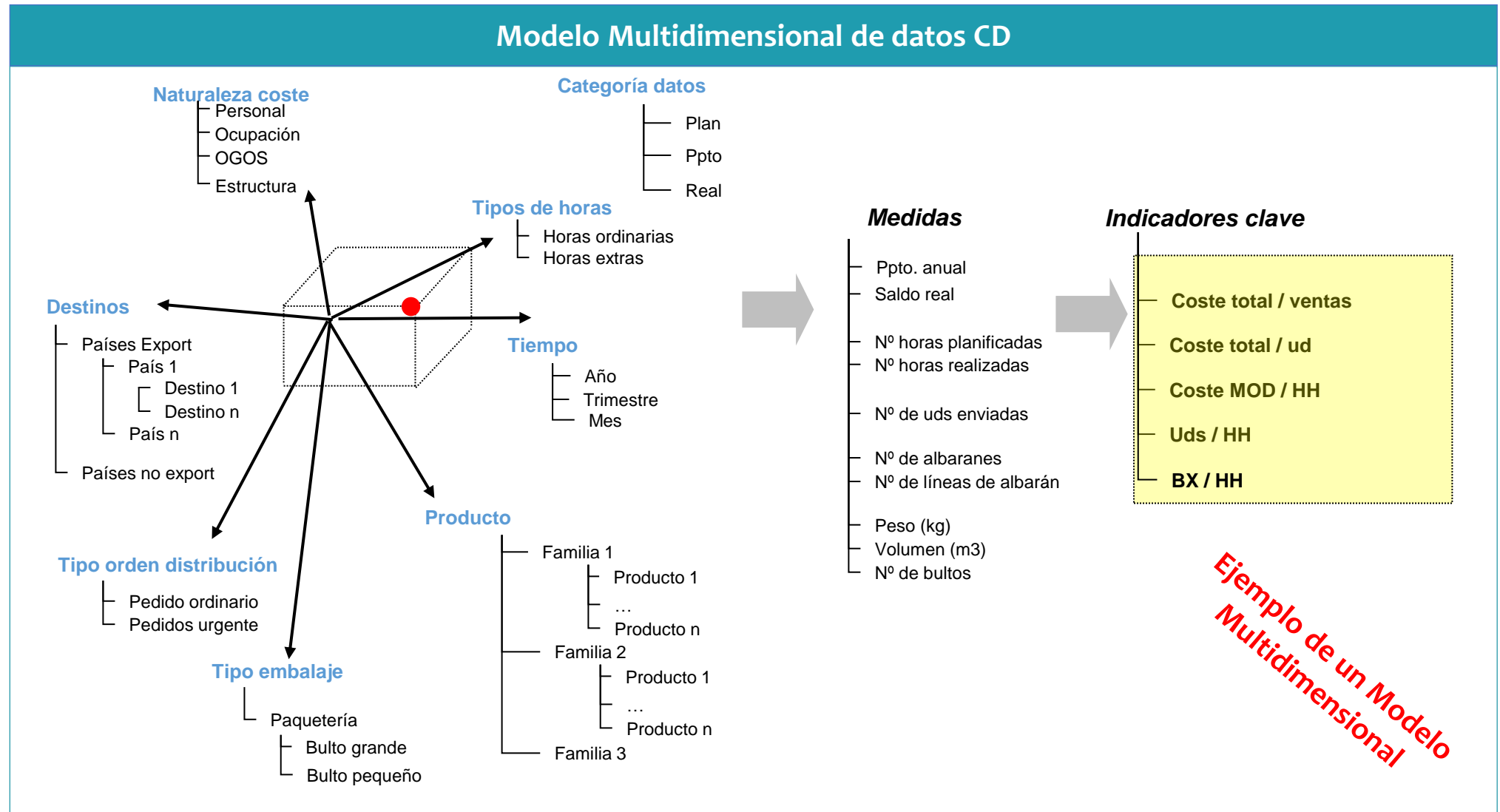
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación



# Modelización Multidimensional | Enfoque

Existen varios enfoques de diseño de un DATAWAREHOUSE

- El enfoque dimensional se refiere al enfoque de **Ralph Kimball** en el que se establece que el almacén de datos se debe modelar utilizando un **modelo dimensional / esquema en estrella**.
- El enfoque normalizado, también llamado modelo 3NF (Tercera forma normal) se refiere al enfoque de **Bill Inmon** en el que se establece que el almacén de datos se debe modelar utilizando un modelo E-R / modelo normalizado.

Las ventajas del modelado multidimensional se basan en la facilidad de uso y entendimiento por parte del usuario final.

→ Plataformas de  
datos modernas

→ Datawarehousing

→ Bases de datos  
relacionales

→ **Modelización  
Multidimensional**

→ Proyecto de  
implantación

# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

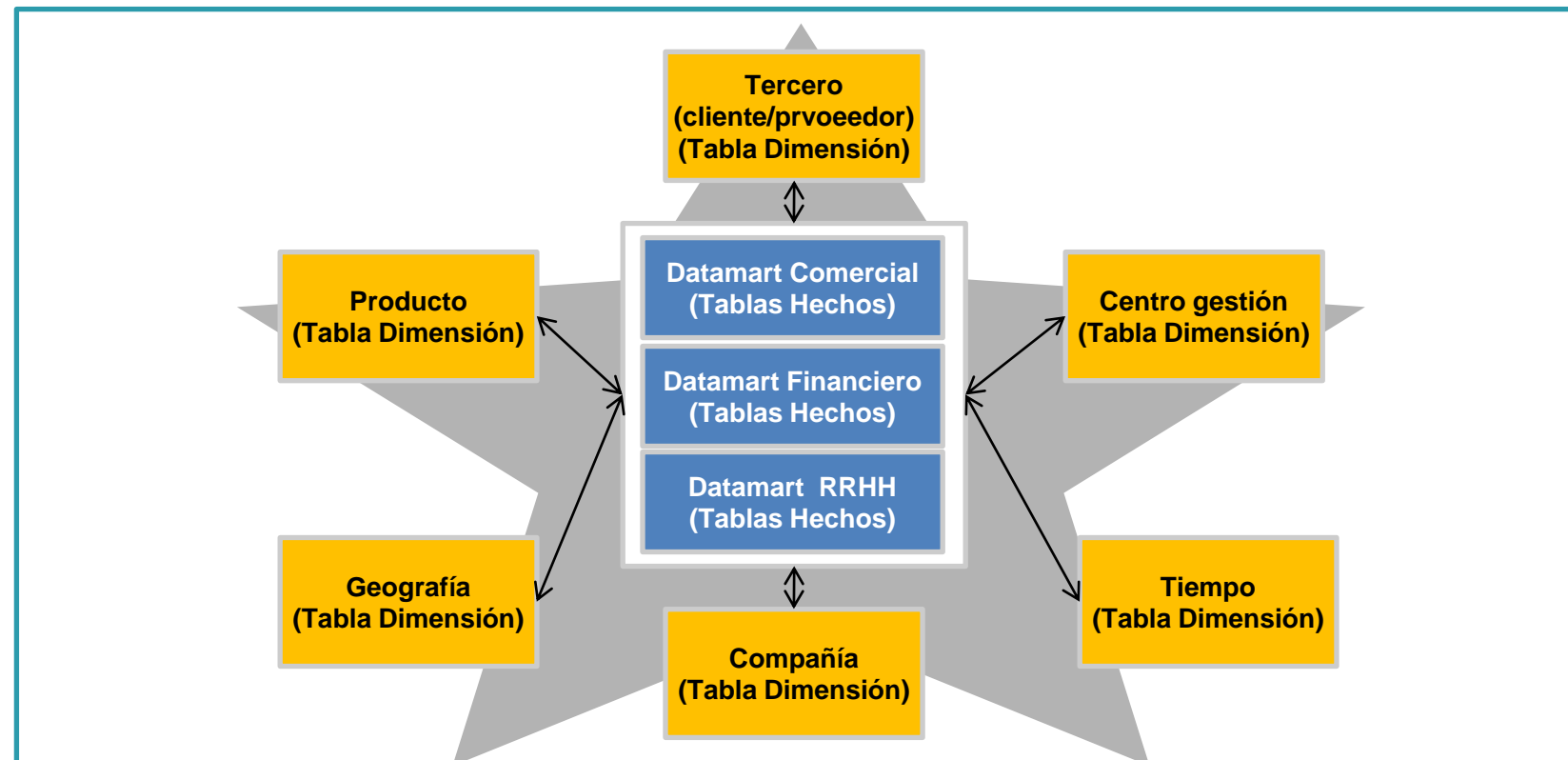
→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

## Modelo de datos en estrella:

- Las **tablas de hechos**, también conocidas como tablas Fact son grandes tablas que almacenan las medidas de los procesos de negocio y típicamente tienen **claves foráneas a las tablas de dimensiones**.
- Las **tablas de dimensiones**, también conocidas como tablas Look-up, contienen atributos descriptivos relativamente estáticos. Tienen claves primarias claramente identificables.



# Modelización Multidimensional

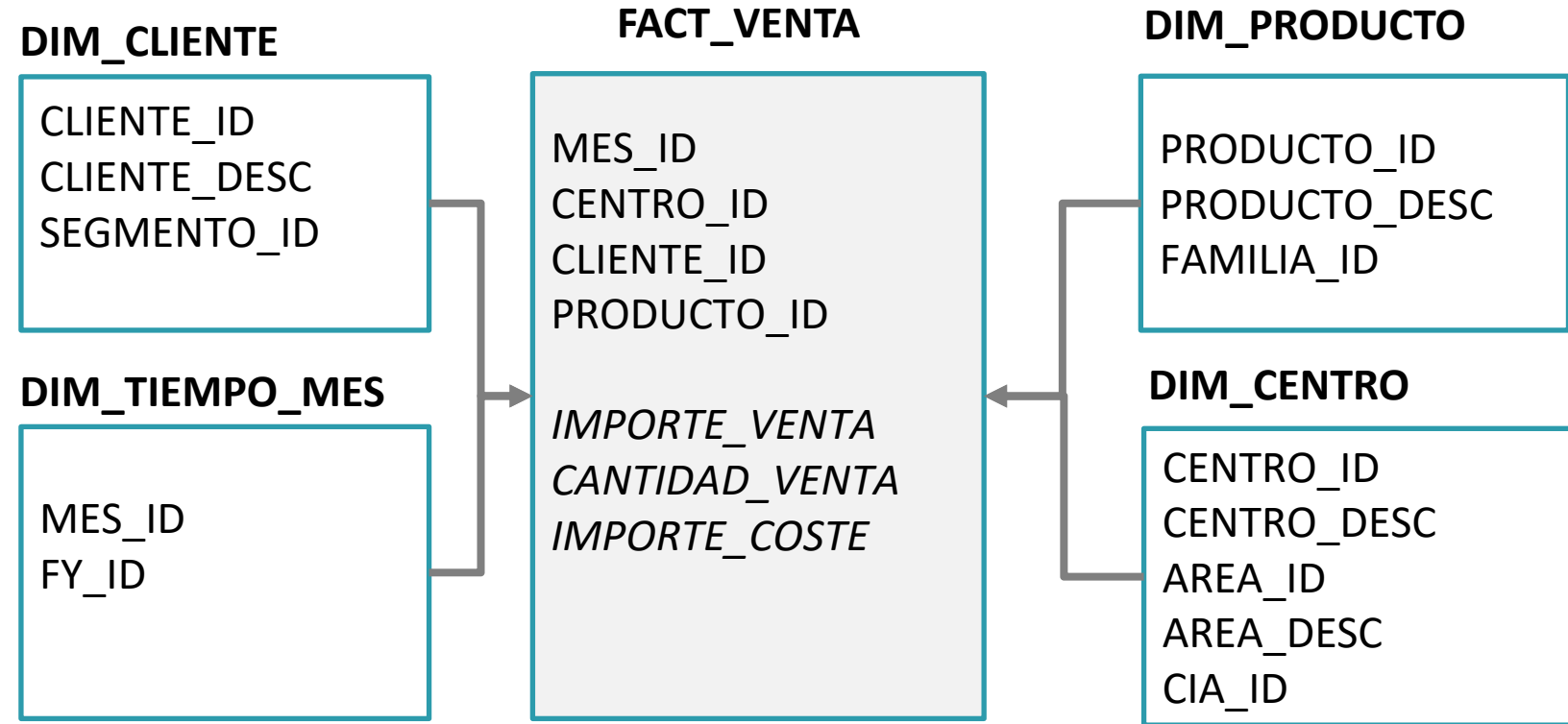
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación



# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

## Proceso para la modelización multidimensional en los DW

- **SELECCIONAR PROCESOS DE NEGOCIO**
  - **Hablar con los usuarios para recoger lo que les gustaría entender utilizando los datos.**
  - Por ejemplo: A los usuarios de Departamento Comercial les gustaría entender las ventas como ayuda para la presentación de informes mensuales y la previsión.
- **DECLARAR LA GRANULARIDAD**
  - **Granularidad define el nivel de detalle que debería estar disponible en el modelo multidimensional**
  - **Idealmente, debería captar el nivel más bajo de detalle posible, que servirá para obtener más flexibilidad en las consultas de los usuarios**
  - Por ejemplo: Seleccionar el resumen de ventas diarias por producto por código postal de cliente.
- **SELECCIONAR LAS DIMENSIONES**
  - Una vez definida la granularidad, se determinará la dimensionalidad primaria de la tabla de hechos
  - **Se pueden agregar nuevas dimensiones a los hechos sin volver a cargar los datos, siempre y cuando no se cambie la granularidad**
  - Por ejemplo: cliente, producto, fecha y código postal son los principales las dimensiones de los hechos.
- **IDENTIFICAR LAS MEDIDAS**
  - **Una vez definida la granularidad y dimensiones de los hechos se han de identificar las medidas necesarias.**
  - Si hay un requisito de medidas no aditivas, tales como porcentajes y ratios, almacenar las medidas subyacentes en el hecho.
  - Por ejemplo: Los usuarios están interesados en el análisis las ventas en unidades e importe y también están interesados en el margen bruto (entonces hay que añadir el coste a los hechos no el margen bruto calculado).

# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

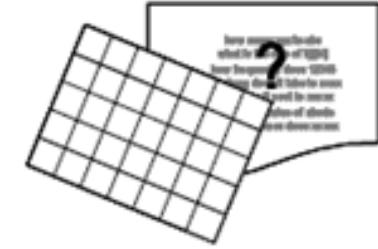
→ **Modelización Multidimensional**

→ Proyecto de implantación

Fases a cubrir hasta la obtención del modelo multidimensional de datos (lógico)

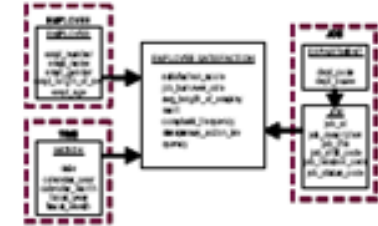
## Análisis de procesos y requisitos de negocio

- Análisis de procesos
- Modelo de requisitos de negocio
- Empezar desde lo vago e incierto
- Evolucionar hasta lo específico y cierto
- **Lista de preguntas de negocio y matriz (hechos y calificadores)**



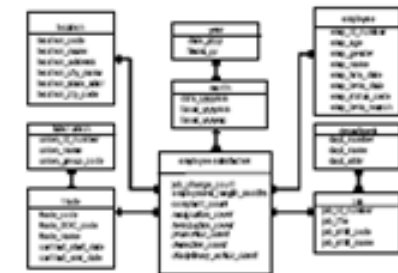
## Modelo lógico de datos

- Diseño (no técnico) conceptual de datos
- Modelo de solución de negocio
- Empezar desde los requisitos de negocio
- Evolucionar hasta las especificaciones de la solución.
- **Modelo dimensional de datos (lógico)**



## Modelo físico de datos

- Implantación (técnica) del modelo diseñado.
- Modelo técnico de la solución.
- Empezar desde la especificación de la solución.
- Evolucionar hasta la especificación de la base de datos.
- **Modelo (físico)**



# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

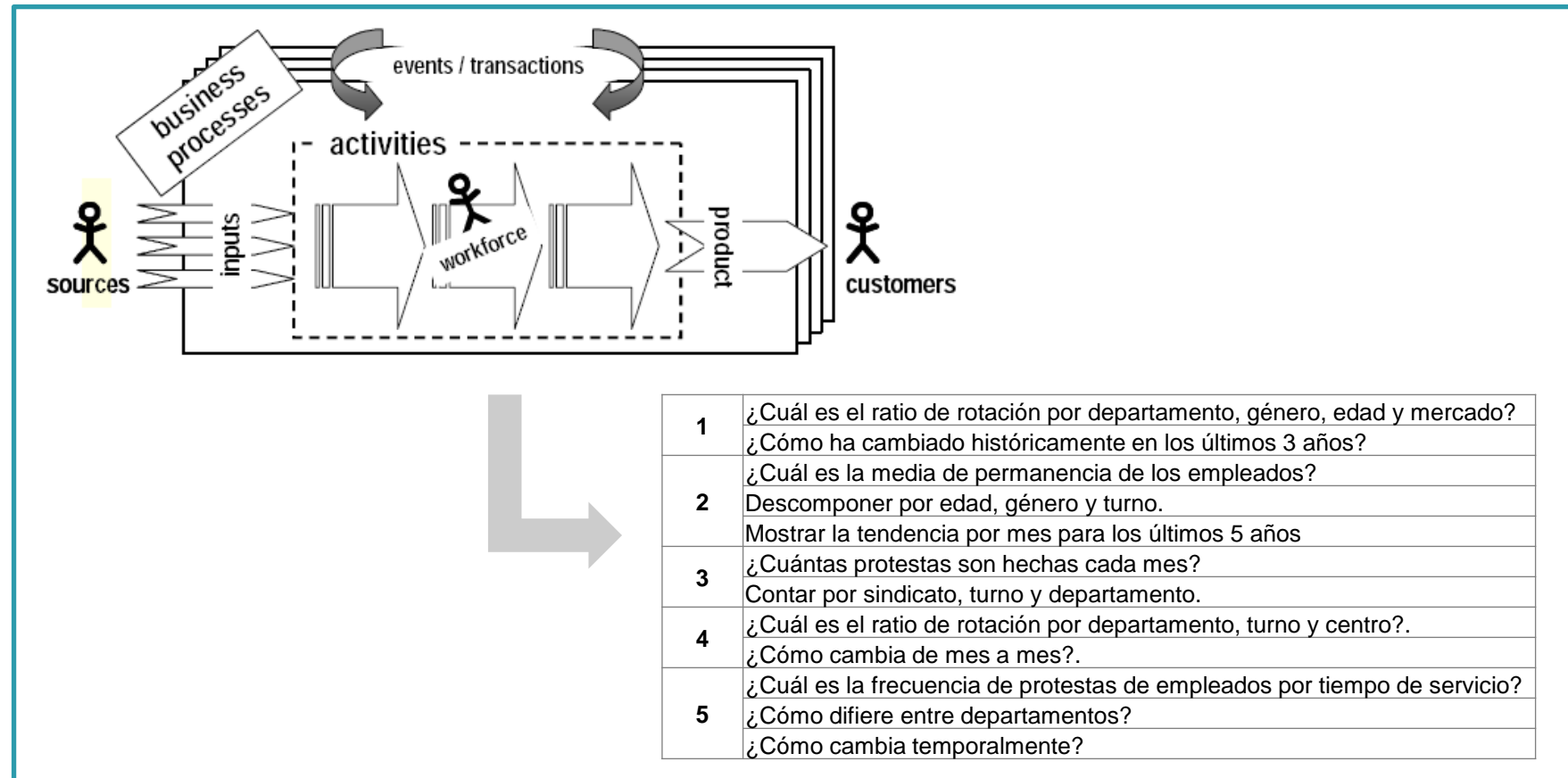
→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

## Recogiendo los Requisitos de Negocio:

- ¿Qué procesos de negocio están en el alcance del modelo?
- ¿Qué puede ser medido sobre los orígenes e inputs?
- ¿Qué puede ser medido sobre las actividades?
- ¿Qué puede ser medido sobre los eventos y transacciones?





# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

## Matriz de Hechos y Cualificadores (F/Q)

- Es necesario **mapear los hechos objeto de análisis** (surgen de las preguntas) con las **dimensiones que los cualifican**.
- En la matriz F/Q se colocan todas las cuestiones de negocio

	ratio de rotación	media permanencia	número protestas					
departamento	1,4		3					
género	1	2						
edad	1	2						
mercado	1							
año	1	2						
turno	4	2	3					
mes	4	2	3					
sindicato			3					
centro	4							

*Hay hechos (p.ej. Ratio de rotación) que aparecen en más de una pregunta.*

*La misma dimensión puede aparecer para más de un hecho.*

# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

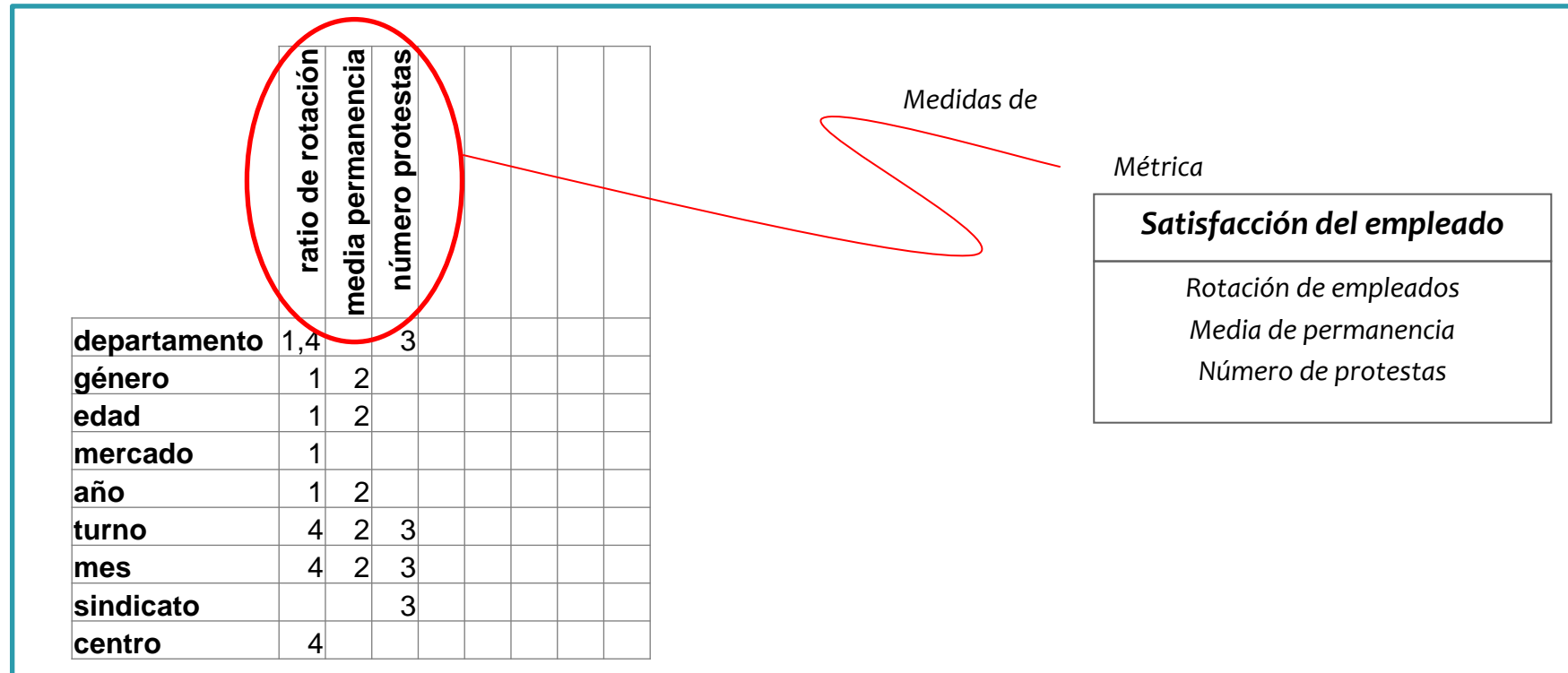
→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

## ○ Definición de la métrica

- El primer paso en la modelización multidimensional es definir la métrica y sus medidas.
- Una métrica es un conjunto de medidas relacionadas que hacen referencia a un concepto de negocio objeto de análisis.
- Por ejemplo: satisfacción del empleado, fidelización de clientes, rentabilidad, eficiencia de la planta,...



# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

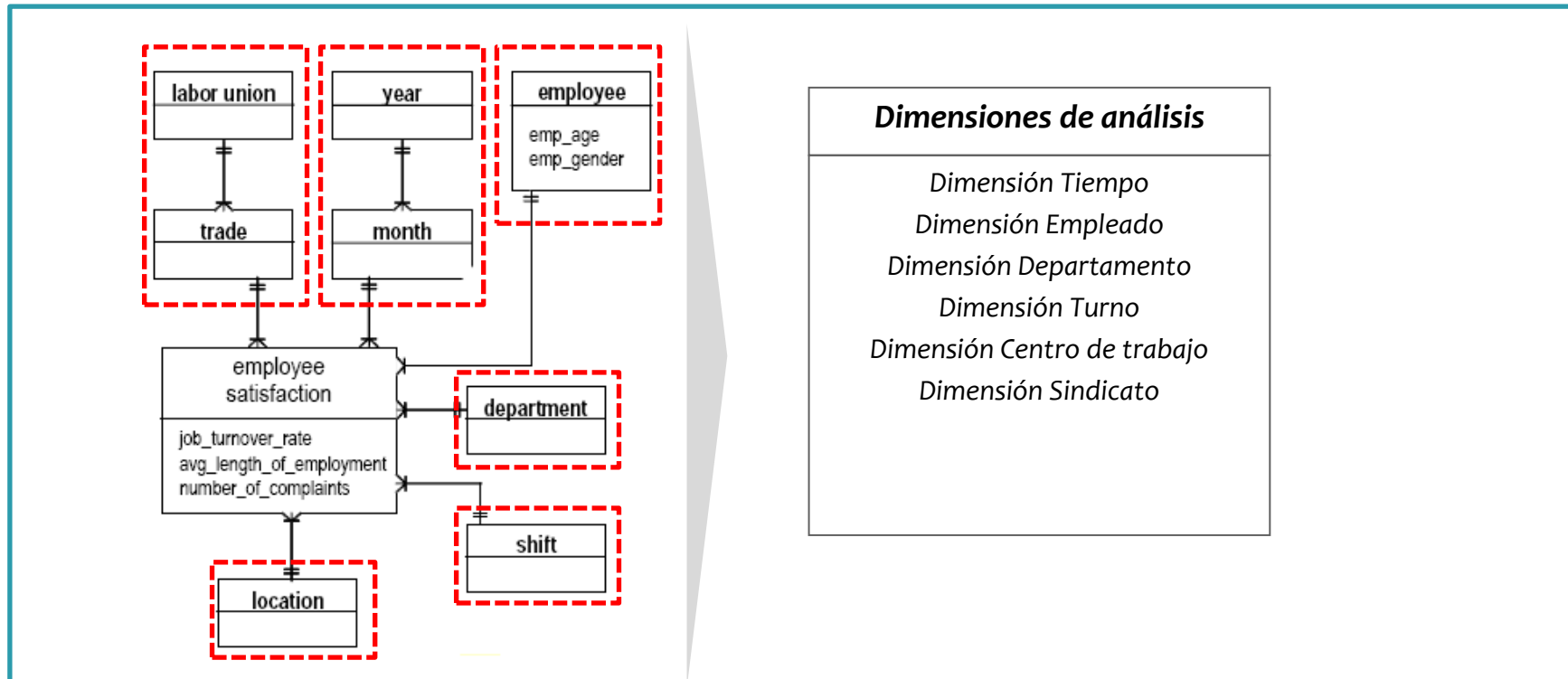
→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

## Asociación de cualificadores a dimensiones:

- Para cada medida contenida en la métrica, todos los cualificadores asociados han de ser representados por una dimensión.
- Identificar los cualificadores que son niveles de una jerarquía.
- Identificar los cualificadores que son atributos de una dimensión



# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

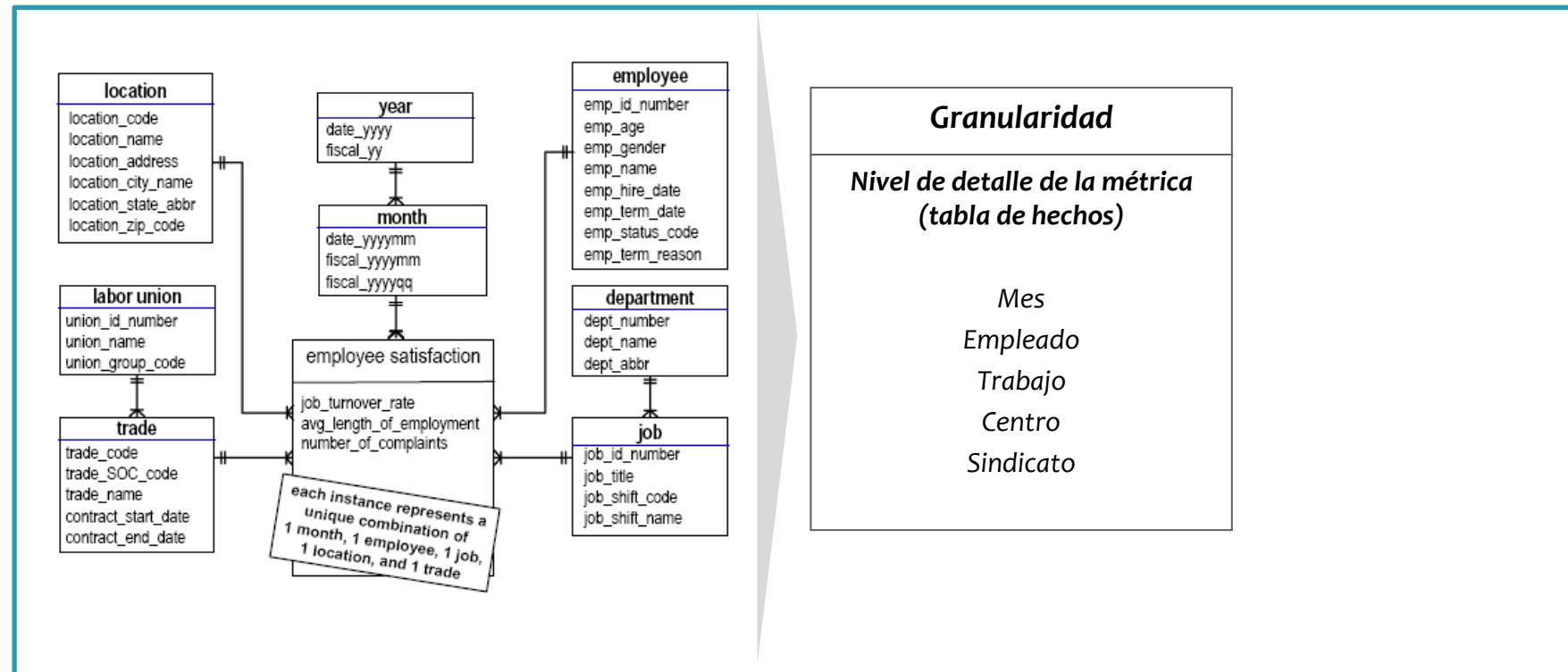
→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

## Definición de la granularidad

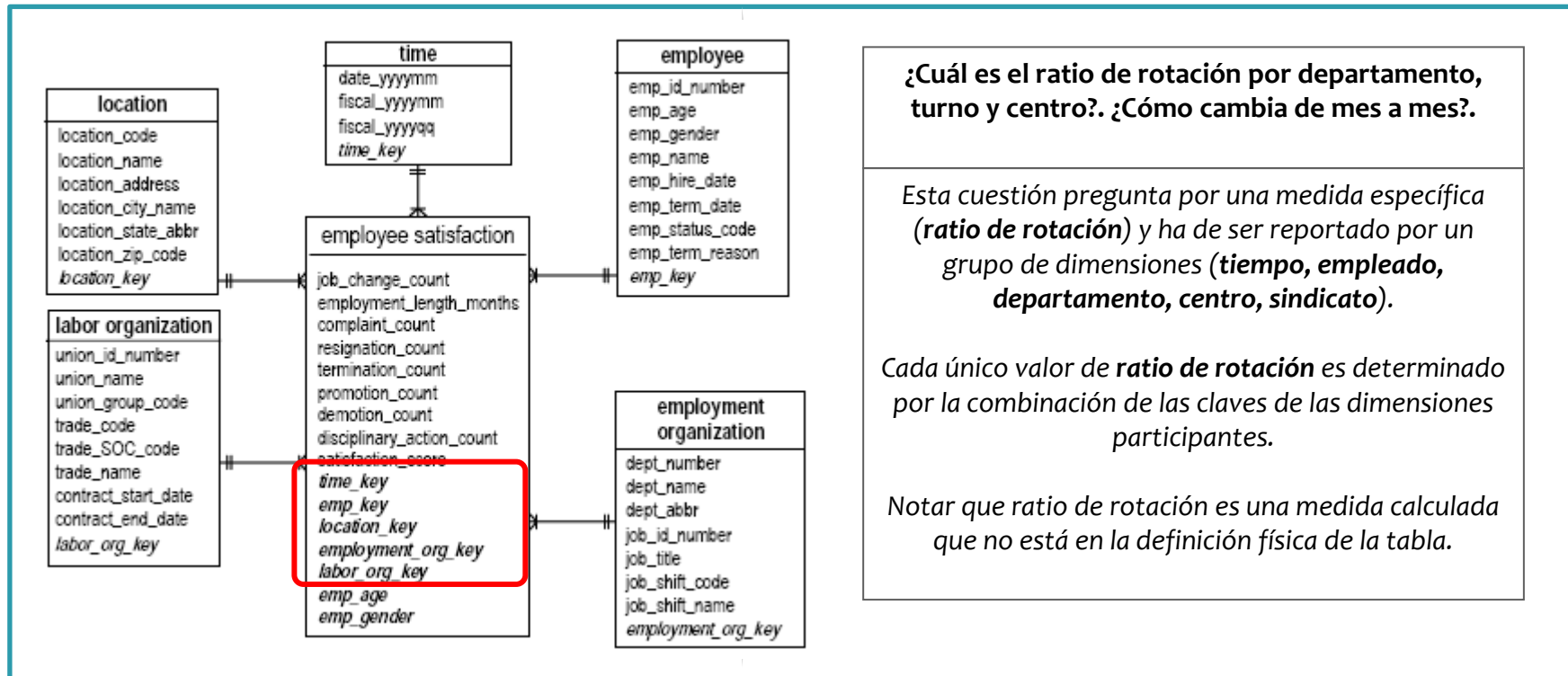
- La granularidad describe el nivel de detalle contenido en la métrica.
- Cada medida en una métrica debe de estar al mismo nivel de detalle.
- La granularidad se determina por el conjunto de dimensiones asociadas con la métrica.



# Modelización Multidimensional

## Definición de la clave de la tabla de hechos:

- El fundamento del concepto OLAP es la navegación desde un conjunto de dimensiones a los hechos asociados con estas dimensiones.
- La clave de la tabla FACT es la combinación de todas los valores de las



# Modelización Multidimensional

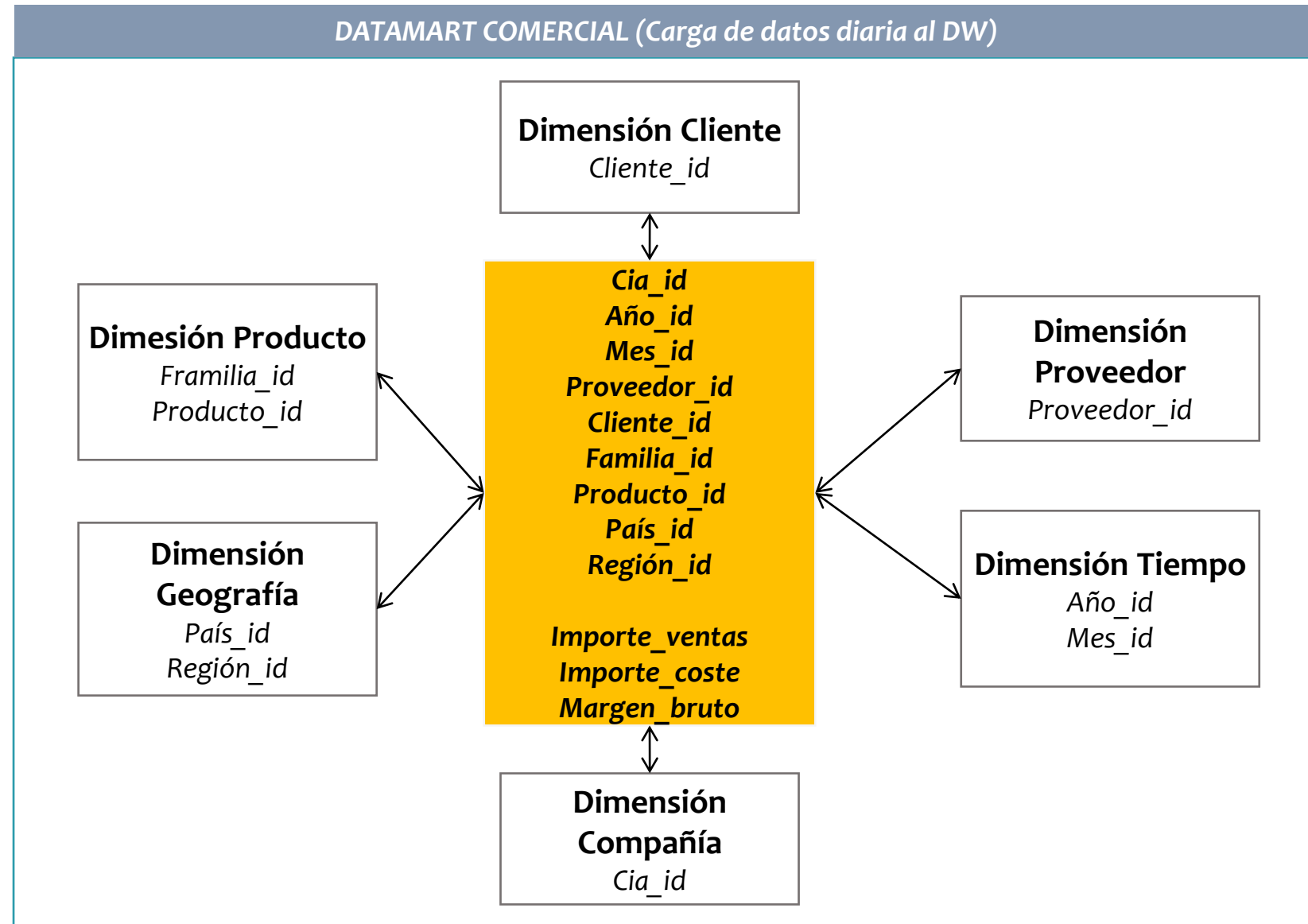
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación



# Modelización Multidimensional

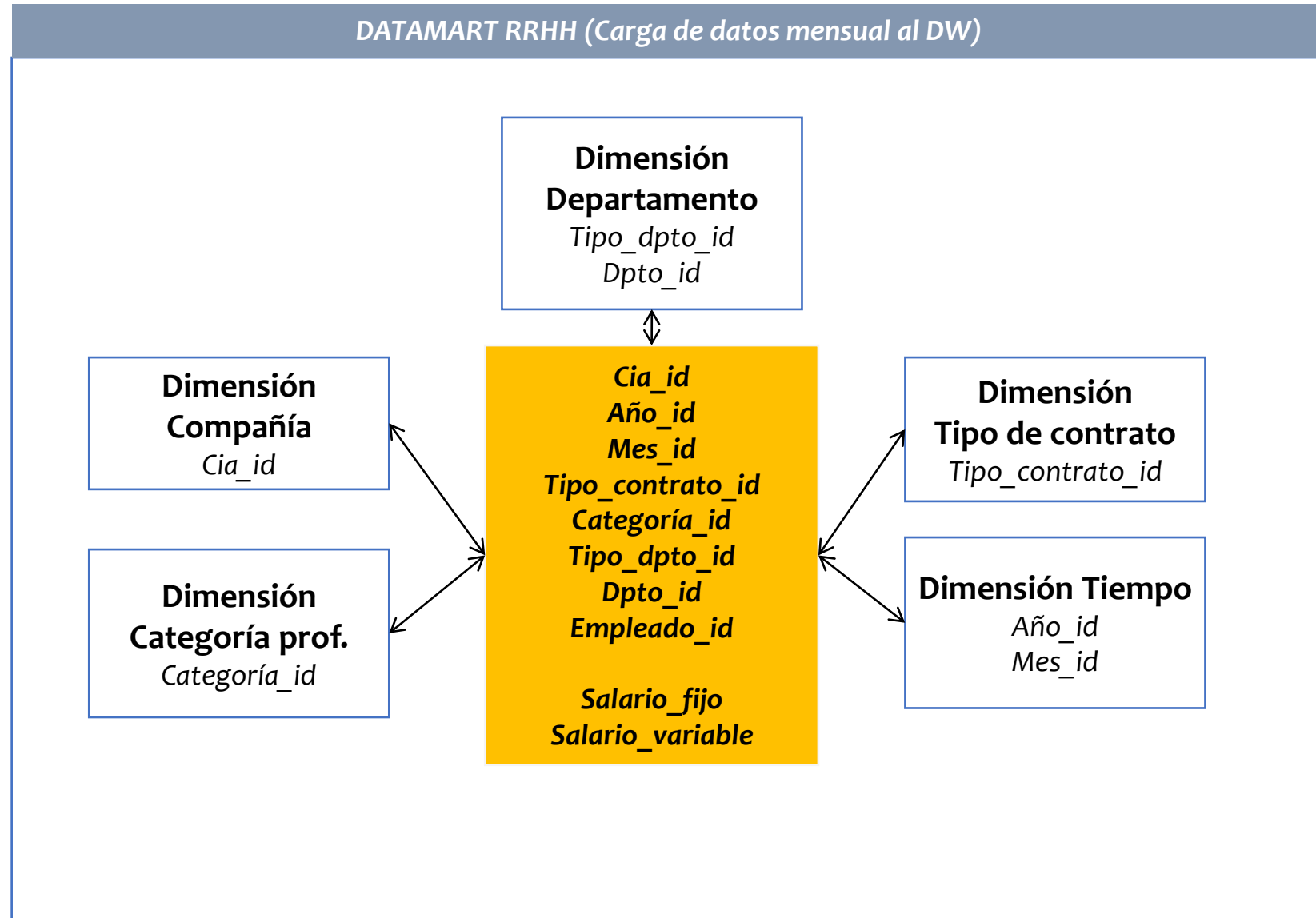
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación



# Modelización Multidimensional

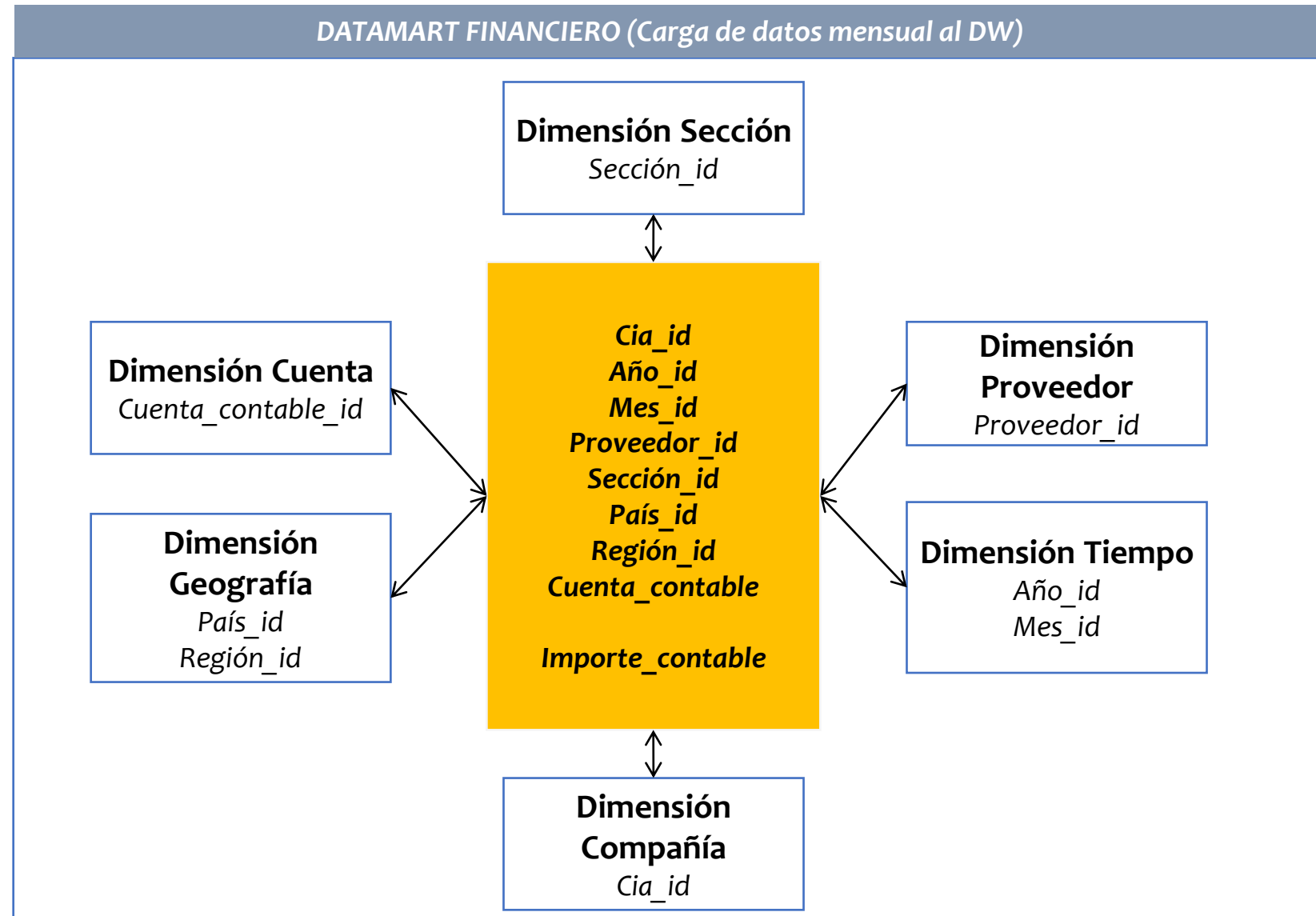
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación





# Modelización Multidimensional

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

- Las **dimensiones** existentes en el modelo son comunes para las diferentes **tablas de hechos**.
- Cada Data Mart utiliza aquellas que son relevantes para su ámbito de análisis.

	<i>Tiempo</i>	<i>Compañía</i>	<i>Cliente</i>	<i>Proveedor</i>	<i>Geografía</i>	<i>Producto</i>	<i>Cuenta</i>	<i>Sección</i>	<i>Departamento</i>	<i>Categoría</i>	<i>Tipo contrato</i>
Datamart Comercial	✓	✓	✓	✓	✓	✓					
Datamart Financiero	✓	✓		✓			✓	✓			
Datamart RRHH	✓	✓							✓	✓	✓

# Modelización Multidimensional | Ejercicio 3

Ejercicio: Seleccionar medidas y dimensiones por área de negocio

Proceso de negocio	Medidas
Ventas y Marketing	
Logística	
Producción	
Recursos Humanos	
TI	

Proceso Negocios	Dimensiones analítica
Ventas y Marketing	
Logística	
Producción	
Recursos Humanos	
TI	

- Plataformas de datos modernas
- Datawarehousing
- Bases de datos relacionales
- **Modelización Multidimensional**
- Proyecto de implantación

# Modelización Multidimensional | Ejercicio 3

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ **Modelización Multidimensional**

→ Proyecto de implantación

Proceso de negocio	Medidas
<b>Ventas y Marketing</b>	Número de Pedidos Comisiones Descuentos Importe Ventas Cantidad Vendida Número de Devoluciones Unidades Vendidas Margen Bruto
<b>Logística</b>	Unidades expedidas Bultos expedidos Palets expedidos Peso Volumen Coste Inventario Unidades recibidas Bultos recibidos Palets recibidos Distancias Costo por Unidad Costo por Kilometro
<b>Producción</b>	Tiempo de Producción Capacidad Usada Unidades Planificadas Unidades Producidas Costos de Producción Horas Hombre
<b>Recursos Humanos</b>	Sueldos y Salarios Pagados Número de Empleados Rotación Absentismo
<b>TI</b>	Proyectos desviados en plazo Coste incurrido Número de Usuarios de BI Número de Usuarios Finales Tiempo en soporte

Proceso Negocios	Dimensiones analítica
<b>Ventas y Marketing</b>	Tiempo Producto Centro Cliente Segmento de Clientes Territorios Formas de Pago
<b>Logística</b>	Tiempo Centro Proveedores Clientes Productos Tipo embalaje Transportistas Tipos de Distribución Puntos de Embarque
<b>Producción</b>	Centro Producto Procesos Estaciones de Trabajo
<b>Recursos Humanos</b>	Tiempo Centro Rangos de Salario Posición (puesto)
<b>TI</b>	Tiempo Centro Proyecto Empleado

# Proyecto de Implantación

Fases

# Proyecto de implantación

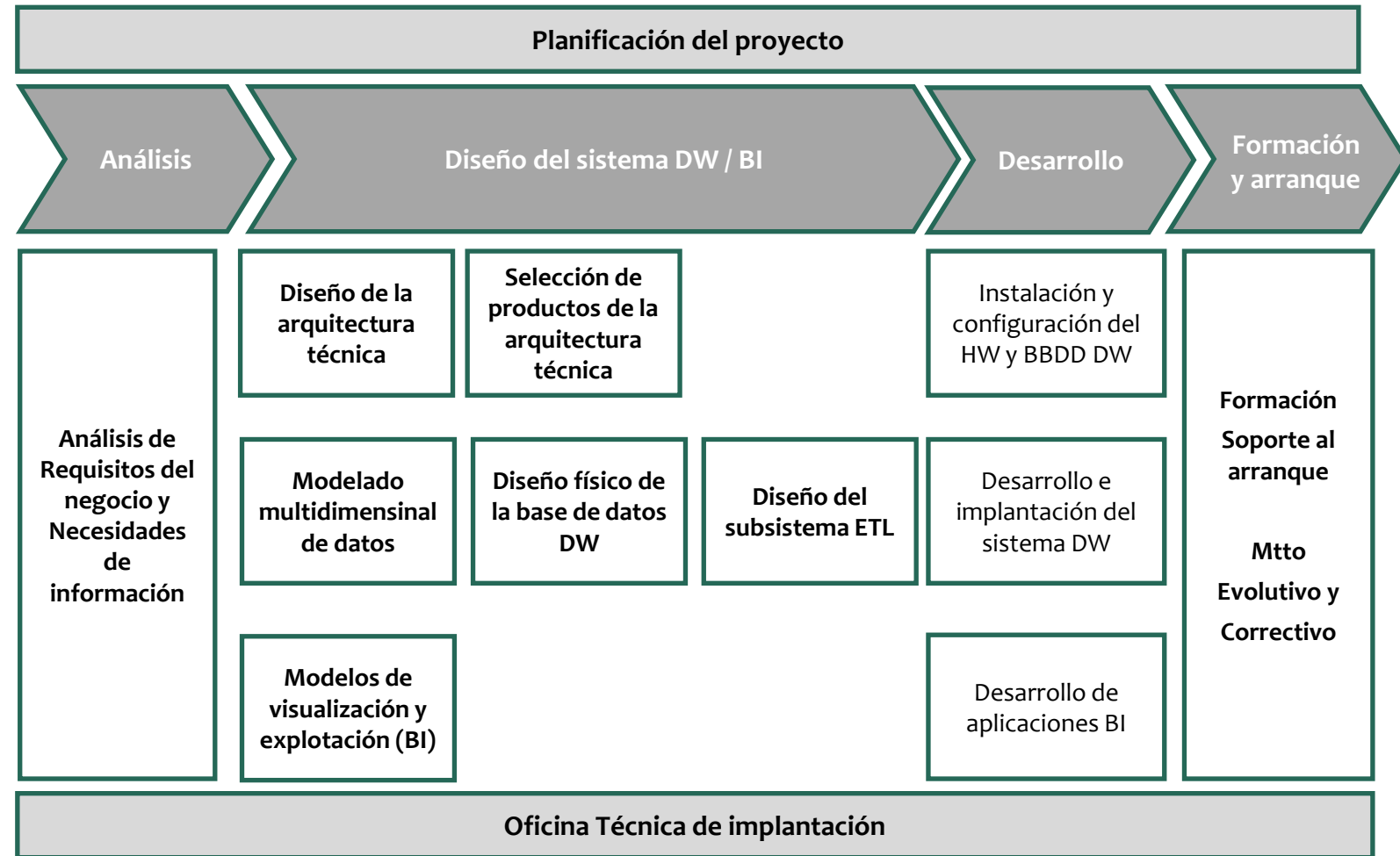
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación



# Proyecto de implantación

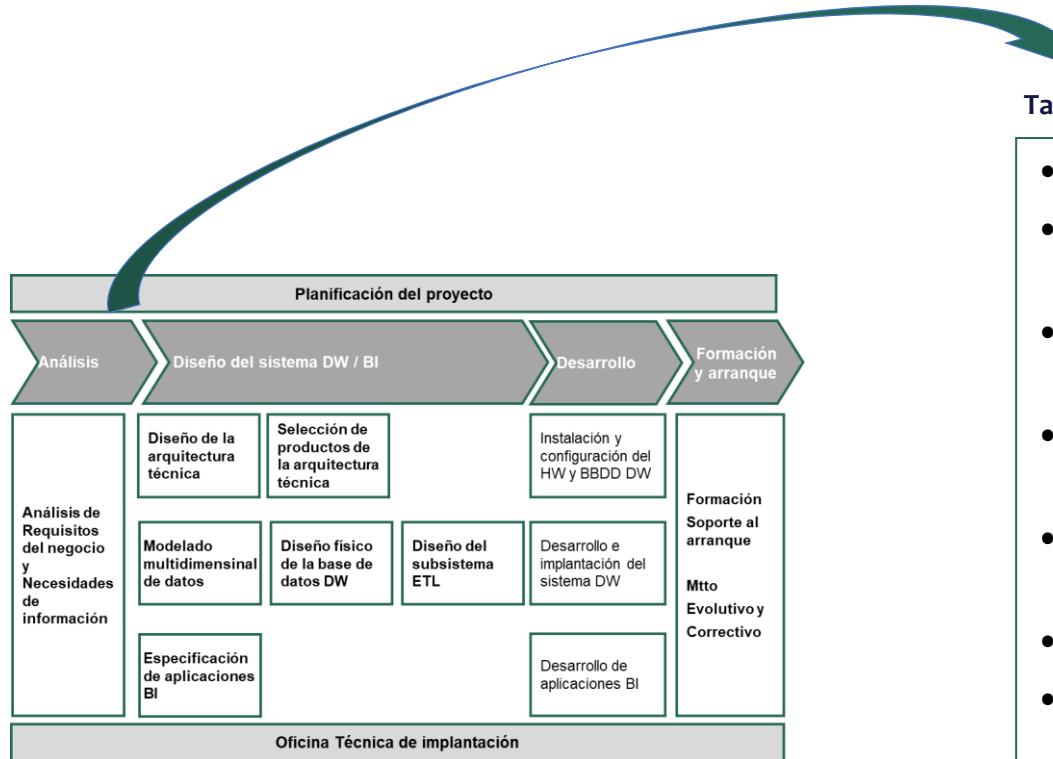
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación



## Tareas

- Análisis de procesos de negocio
- Análisis del mapa de sistemas transaccionales
- Inventario de informes (origen, frecuencia y destinatarios)
- Identificación de métricas e indicadores clave
- Identificación de dimensiones y ejes del análisis
- Granularidad del análisis
- Análisis de volumetría

## Metodología

- Entrevistas y workshops con las diferentes áreas y departamentos del negocio
- Inventario de informes

# Proyecto de implantación

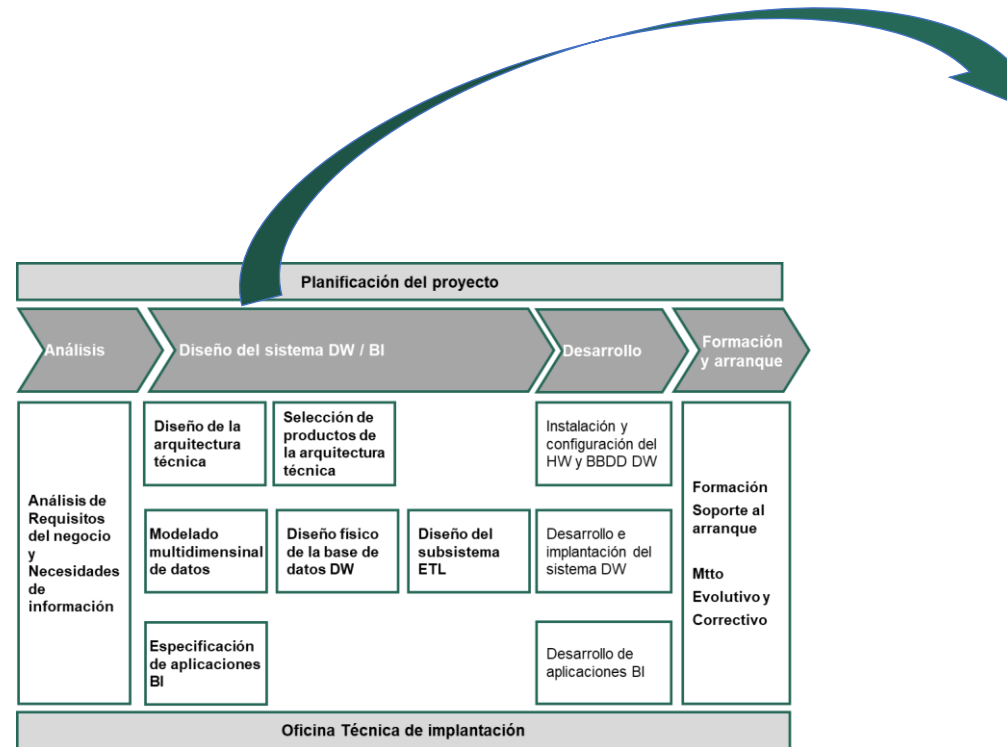
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ **Proyecto de implantación**



## Tareas

- Diseño del Mapa de Sistemas y Arquitectura funcional
- Diseño del Modelo multidimensional de datos (métricas, dimensiones, granularidad)
- Diseño físico de la Base de Datos (tablas de hechos, dimensiones, agregadas, etc.)
- Colaboración con el partner técnico para la especificación de procesos ETL (orígenes de datos y transformaciones)
- Definición de la seguridad (roles y usuarios)
- Diseño de los modelos de visualización y explotación (BI)

## Metodología

- Workshops con el área de TI
- Workshops con las áreas de negocio
- Metodología de modelado multidimensional

# Proyecto de implantación

→ Plataformas de  
datos modernas

→ Datawarehousing

→ Bases de datos  
relacionales

→ Modelización  
Multidimensional

→ Proyecto de  
implantación

## Aspectos metodológicos claves en la fase de Diseño

- **Diseño de la arquitectura técnica:** en esta fase se deben tener en cuenta tres factores: los requerimientos de negocio, los actuales entornos técnicos, y las directrices técnicas y estratégicas futuras planificadas por la compañía, lo que permitirá establecer el diseño de la arquitectura técnica del entorno del Data Warehouse.
- **Modelado Dimensional:** se comienza con una matriz donde se determina la dimensionalidad de cada indicador para luego especificar los diferentes grados de detalle dentro de cada concepto del negocio.
- **Diseño Físico:** se centra en la selección de las estructuras necesarias para soportar el diseño lógico. Un elemento principal de este proceso es la definición de estándares del entorno de la base de datos. La indexación y las estrategias de particionamiento se determinan en esta etapa.
- **Diseño de la capa de integración:** tiene como principales actividades la extracción, transformación y carga (ETL). Estas actividades son altamente críticas ya que tienen que ver con la identificación de los orígenes de datos, así como lo definición de los procesos de transformación necesarios.
- **Definición del modelo de reporting:** Tiene que ver con el modo en el que se explotarán los datos. Lo cual incluye la definición de tres tipos de herramientas: cuadros de mando de alto nivel, modelos analíticos y sets de reporting (desatendidos y/o a demanda).
- **Definición de la seguridad:** Identificación de los diferentes roles y usuarios con los diferentes niveles de acceso en función del nivel jerárquico y área funcional.



# Proyecto de implantación

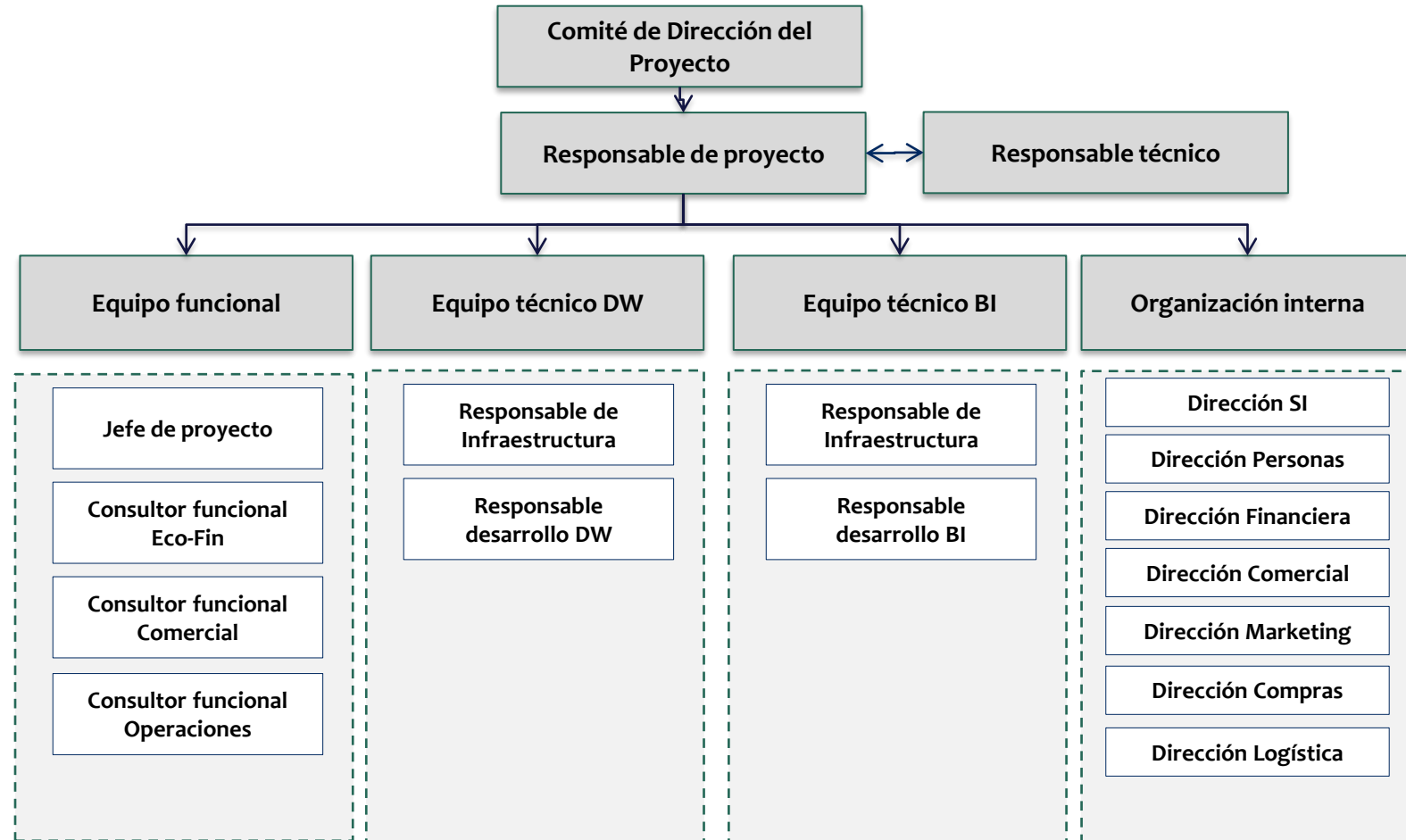
→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación



- Oficina Técnica del Proyecto
- Liderar el Diseño Conceptual y Funcional
- Supervisar el Desarrollo Técnico.
- Transferencia de conocimiento a equipos internos

- Desarrollo DW (integración, modelos de datos, modelos analíticos)
- Instalación infraestructura
- Pruebas

- Desarrollo BI
- Instalación infraestructura
- Pruebas

# Proyecto de implantación

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

→ Proyecto de implantación

Los roles de los distintos actores del proyecto:

<b>Comité de Dirección del Proyecto</b>	<ul style="list-style-type: none"> <li>• Órgano de gobierno del proyecto</li> <li>• Compuesto por altos directivos</li> <li>• Aprobación diseños y entregables</li> <li>• Reunión mensuales o a demanda</li> </ul>
<b>Comité de Seguimiento del Proyecto</b>	<ul style="list-style-type: none"> <li>• Compuesto por Usuarios clave y Dirección y consultores</li> <li>• Realizar la toma de decisión de los aspectos tácticos y técnicos del proyecto.</li> <li>• Es posible, que ciertas decisiones que tome este órgano deban ser validadas por el Comité de Dirección.</li> </ul>
<b>Clientes internos</b>	<ul style="list-style-type: none"> <li>• Los clientes internos principales del proyecto son todos los departamentos y áreas de negocio de la organización</li> <li>• Los clientes internos del proyecto serán informados a lo largo del mismo, y podrán ser consultados y/o participes en ciertas tomas de decisiones.</li> </ul>

# Proyecto de implantación

→ Plataformas de datos modernas

→ Datawarehousing

→ Bases de datos relacionales

→ Modelización Multidimensional

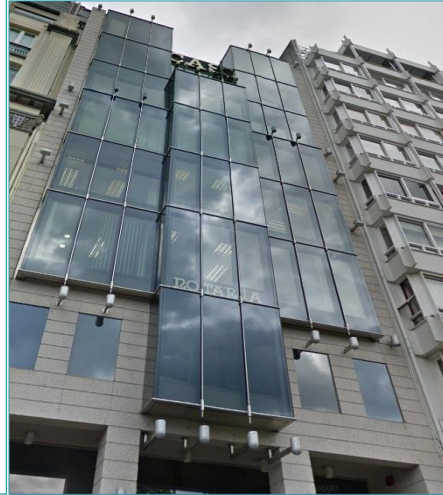
→ Proyecto de implantación

<b>Técnico administrador de infraestructura</b>	<ul style="list-style-type: none"> <li>• Responsable de instalación de la solución</li> <li>• Asegurar que los servicios están corriendo y trabajando perfectamente</li> <li>• Seguridad y Back-ups de los datos</li> </ul>
<b>Técnico administrador de base de datos</b>	<ul style="list-style-type: none"> <li>• Responsable de la carga de datos desde los orígenes transaccionales (conocimiento de los orígenes de datos). Control de la carga de datos</li> <li>• Responsable de la creación de nuevos objetos/queries/informes en la plataforma</li> </ul>
<b>Controller responsable de la calidad del dato</b>	<ul style="list-style-type: none"> <li>• Profundo entendimiento del negocio. Lidera la interpretación de los requisitos de información del negocio</li> <li>• Responsable de la calidad de los datos. Control de los mappings de los datos maestros (Cuentas, CEBEs)</li> <li>• Mejora de procesos para que tanto tablas de hechos como maestros estén perfectamente informados</li> <li>• Filtrar los requisitos de información de los usuarios de negocio</li> </ul>
<b>Usuario de negocio (perfil ejecutivo)</b>	<ul style="list-style-type: none"> <li>• Dirección general, Direcciones de área, Segunda línea de dirección</li> <li>• Es un perfil que analiza a alto nivel los datos que se le muestran pero no entra al detalle</li> <li>• Normalmente será usuario de Cuadro de Mando (Dashboards) y recibirá informes pdf</li> </ul>
<b>Usuario de negocio (perfil analista)</b>	<ul style="list-style-type: none"> <li>• Igual que el anterior, pero adicionalmente (por su perfil) necesitará herramientas de análisis para profundizar en los datos</li> <li>• Habrá de conocer perfectamente el modelo de datos del sistema</li> </ul>
<b>Usuario de negocio (perfil consumidor)</b>	<ul style="list-style-type: none"> <li>• Sólo consume informes pdf predeterminados en ciclos diario, semanal y mensual</li> <li>• Para mayor profundización en los datos tendrá los orígenes transaccionales de los datos</li> </ul>

# Contacto

- Plataformas de datos modernas
- Datawarehousing
- Bases de datos relacionales
- Modelización Multidimensional
- Proyecto de implantación

**LA CORUÑA**  
**(Headquarters)**  
 Cantón Pequeño 13, 5º  
  
 Phone: 881 878 263  
 Email: [info@hintd.net](mailto:info@hintd.net)  
  
[www.hintd.net](http://www.hintd.net)



**MADRID**  
 Paseo de la Castellana 93, 2ª



Javier Picos García  
**PARTNER**

- ☎ 619 285 623
- ✉ [javier.picos@hintd.net](mailto:javier.picos@hintd.net)
- in [linkedin.com/in/javierpicos](https://www.linkedin.com/in/javierpicos)
- 📍 Cantón Pequeño 13, 5º B-T. 881 878 263

Corporate Finance | Advanced Analytics | Performance Management