



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE CIENCIAS
FUNDAMENTOS DE BASES DE DATOS

Conceptos básicos

Gerardo Avilés Rosas
gar@ciencias.unam.mx

¿Por qué son importantes?

Las **bases de datos** son el método preferido para el almacenamiento **estructurado de datos**: ya sea que estés navegando a través de fotos y/o publicaciones de **Facebook** de alguno de tus amigos o compartiendo imágenes en **Pinterest** o “twitteando” acerca de cuánto amas tu unicorn frappuccino en **Twitter**, estás continuamente viendo, manipulando y creando datos.





...¿Por qué son importantes?

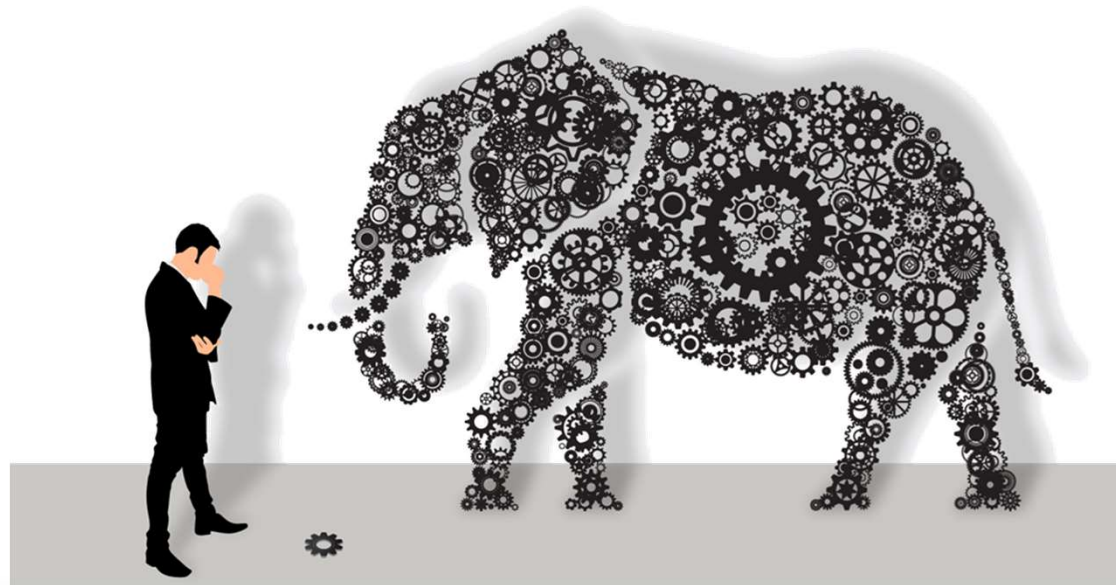
Incluso, toda la magia que sucede en la **Web**, son realmente una serie de **aplicaciones de bases de datos**: nuestras interacciones con sitios web están restringidas por las limitaciones inherentes de bases de datos (cada byte tiene que ser almacenada en lo que podría describirse como una base de datos colosal).



Todos utilizan tecnología de bases de datos para asegurar la **integridad** de los datos y **facilitar** la labor, tanto de los usuarios como de los programadores que los desarrollan.

...¿Por qué son importantes?

El **reto más importante** en la aplicación de una base de datos, es el **diseño correcto** de su estructura. Sin una comprensión completa del problema que la base de datos pretende resolver, y sin conocimiento de las mejores prácticas para la organización de los datos requeridos, la base de datos se convierte en un animal difícil de manejar que requiere atención constante.





¿Por qué usar una Bases de Datos?

Juan Pérez tiene muchos clientes y tiene una gran cantidad de información acerca de ellos en notas garabateadas como ésta:



...¿Por qué usar una Bases de Datos?

Desafortunadamente **Juan** a usado este sistema durante mucho tiempo. La semana pasada expandió sus contactos para incluir en sus lista a personas que están buscando trabajo, de manera que su lista creció rápido, bastante rápido...



...¿Por qué usar una Bases de Datos?



¿Será la mejor
forma de
organizar y
almacenar mi
información?



...¿Por qué usar una Bases de Datos?

Obviamente es una mala idea la forma de organización que eligió Juan. **¿Qué podemos hacer?**



...¿Por qué usar una Bases de Datos?

Como se puede observar, hay información similar almacenada en cada nota:



Nombre

Apellidos

Ocupación

Fecha de nacimiento

Dirección

Email

Estado

...¿Por qué usar una Bases de Datos?

De acuerdo a lo anterior, organicemos la información de Juan:

Nombre	Cumpleaños	Ocupación	Edo Civil	Dirección	E-mail	Estado
Carlos Rueda	24 noviembre	Programador	Soltero	Tulipán 34	carlos@h.mx	Cd. Mx.
Laura Salinas	3 de marzo	Maestra	Soltera	Volcán 321	laura@m.mx	Chihuahua
Karla Santos	15 de mayo	Diseñadora	Casada	Frontera 23	Karla@sf.jh	Chiapas
César Rivas	23 octubre	Abogado	Soltero	Bugambilia 12	CesarR@ht.g	Morelos

...¿Por qué usar una Bases de Datos?

¿Qué pasaría si el número de contactos fueran **100, 500, 1,000, 100,000**, etc.?





Limitaciones de MS Excel

- **Excel no llega ni al infinito ni más allá**

- ☐ **Libros abiertos:** Limitado por los recursos disponibles del sistema y memoria
- ☐ **Tamaño de hoja de cálculo:** 1,048,576 filas por 16,384 columnas
- ☐ **Total de caracteres que puede contener una celda:** 32,767 caracteres
- ☐ **Hojas en un libro:** Limitado por la memoria disponible
- ☐ **Campos en un formulario de datos:** 32
- ☐ **Parámetros de un libro:** 255 parámetros por libro
- ☐ **Tamaño de archivo:** en entorno de 32 bits está sujeto a 2 GB y en entorno de 64 bits no se imponen sin límites.

- Para obtener más información:

<https://support.office.com/es-es/article/Especificaciones-y-l%C3%ADmites-de-Excel-1672b34d-7043-467e-8e27-269d656771c3>





Limitaciones de MS Access

- Tamaño del archivo: **2 GB**, menos el espacio necesario para almacenar los objetos que se hayan creado.
- Número de objetos en la base de datos: **32,768**
- Número de usuarios simultáneos: **255**
- Número de campos en una tabla: **255**
- Número de tablas abiertas: **2,048**
- Número de relaciones: **32 por tabla**.



- Para obtener más información:

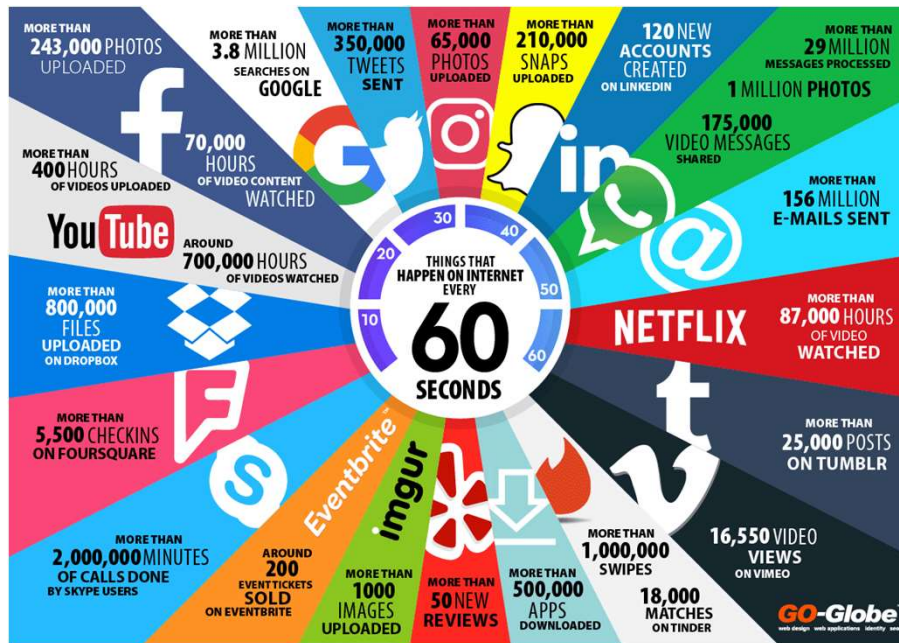
<https://support.office.com/es-es/article/Especificaciones-de-Access-2010-1e521481-7f9a-46f7-8ed9-ea9dff1fa854>



Datos, datos, datos everywhere

El mundo está lleno de datos y mucho de los que hacemos diariamente se convierte en datos para diferentes organizaciones:

Usar cajeros automáticos, obtener una licencia, comprar un libro, hacer una reservación, bajar música, comprar boletos de cine, etc.



¿Qué queremos hacer con ellos?



- Almacenarlos
- Consultarlos
- Actualizarlos

...Datos, datos, datos everywhere

Estamos **ahogados en datos**, pero **carecemos de información**...



Supongamos que requerimos hacer una aplicación para almacenar y/u obtener información sobre multas de tránsito, ¿qué requerimos?

- **Personas**
- **Autos**
- **Oficiales de tránsito**
- **Multas**

Nos interesa saber:

- ☐ ¿De qué persona se trata?
- ☐ ¿Qué auto posee una persona?
- ☐ ¿Qué infracción cometió?
- ☐ ¿Qué oficial levantó la infracción?
- ☐ ¿En dónde ocurrió el hecho?
- ☐ Etcétera.



¿Qué necesitamos?

- **Almacenar** los datos por **mucho tiempo**.
- **Grandes volúmenes** de datos (cientos de **GB**)
- **Proteger** los datos contra desastres.
- **Proteger** los datos contra usos **no autorizados**.
- **Consultar / actualizar:**
 - ❑ ¿Cuántas multas ha cometido una persona?
 - ❑ ¿Cuál es la zona donde hay más percances?
- **Permitir el acceso** a cientos o miles de usuarios.
- **Permitir** modificar el esquema.



- **El almacenamiento de datos:**

- ❑ El sistema de archivos limita el tamaño de éstos:
p.e. en Excel tenemos 1,048,576 filas por 16,384 columnas
- ❑ Cuando el sistema se cae, los datos pueden perderse.
- ❑ La autorización basada en contraseñas es insuficiente.

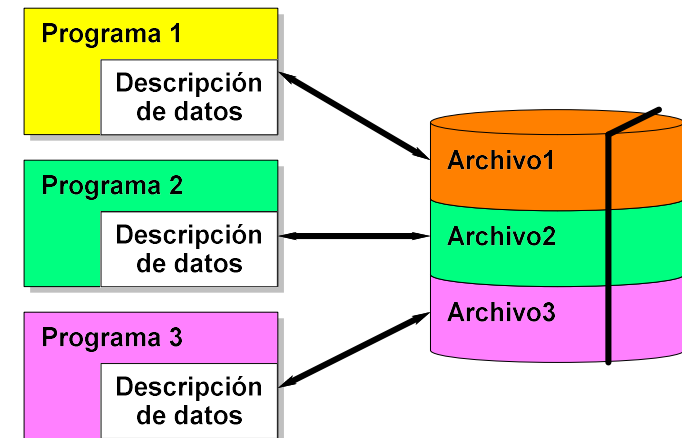
- **Consulta / actualización:**

- ❑ Se necesita escribir un nuevo programa para cada nueva consulta.
- ❑ Es necesario implementar políticas para lograr un buen rendimiento.

- **La concurrencia proporciona una protección limitada:**

- ❑ Necesitas preocuparte que los usuarios no interfieran entre sí.
- ❑ No puedes proporcionar diferentes vistas de los mismos datos.

- **Al cambiar el esquema se requiere cambiar casi todas las aplicaciones.**



- **Datos:**

- ☐ Hechos sin evaluar, que pueden registrarse y que no tienen significado explícito.
- ☐ Hechos relacionados con objetos, eventos o personas, que pueden ser registrados y almacenados en medios informáticos.
- ☐ Unidad mínima de información, valor sin significado explícito.

- **Información:**

- ☐ Conjunto de datos que han sido procesados, de tal forma que el “conocimiento” de la persona que los utiliza se ha incrementado. Poner datos en contexto.
- ☐ Conjunto de datos interrelacionado que tiene un significado del cual podemos obtener “conocimiento” para la toma de decisiones.

- **Conocimiento.** Información adquirida por una persona a través de la experiencia, la educación, la comprensión teórica o práctica de un asunto referente a la realidad.



- Los **datos** constituyen los **bloques de información**.
- La **información** se produce al **procesar los datos** y se utiliza para **revelar el significado** de los datos.
- La información **precisa, relevante y oportuna** es la base para tomar **buenas decisiones**.
- Las **buenas decisiones** son la clave para la **supervivencia** de las organizaciones en un entorno globalizado.
- La **información oportuna y útil** requiere **datos precisos**.
- Los **datos precisos** deben **generarse y almacenarse** en un **formato** que **facilite su acceso** y su **procesamiento**.
- El **Manejo de Datos** es una disciplina que se enfoca en la **generación, almacenamiento y recuperación de datos**.
- El manejar datos, es una actividad importante para cualquier organización.



▪ Base de datos (BD):

- ❑ *Colección de datos relacionados con un objetivo común, que existen durante un largo período de tiempo, que representa aspectos del mundo real; diseñada, construida y poblada con datos para un propósito específico.*
- ❑ *Colección de datos que están lógicamente relacionados entre sí, tienen una definición y descripción comunes, estructurados de forma particular.*
- ❑ *Es un conjunto exhaustivo de datos estructurados, fiables y homogéneos, organizados independientemente de su utilización y de su implementación física, accesibles en tiempo real, compartibles por usuarios concurrentes que tienen necesidades de información diferentes y no predecibles en el tiempo.*

En el lenguaje coloquial, con la expresión base de datos se designa una colección de datos que es administrada por un sistema manejador de bases de datos SMBD (**Data Base Management System**).



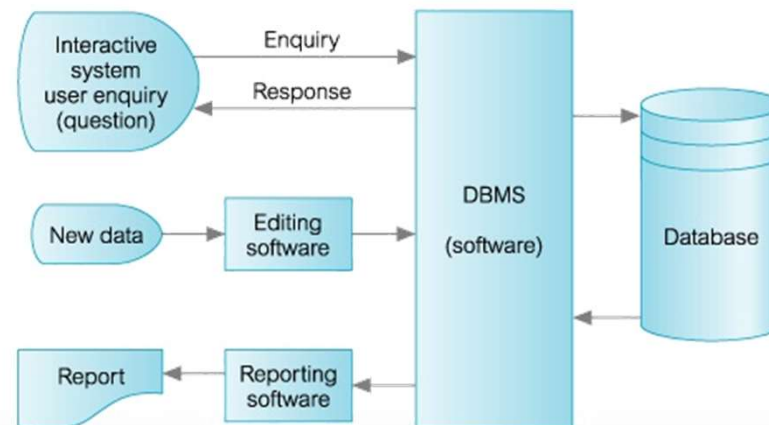
- **Metadatos.** Datos que describen las propiedades o características de los datos del usuario y el contexto de los mismos. Datos acerca de los datos: *nombre de los datos, definición, tamaño o longitud, valores permitidos.*
- **Diccionario de datos.** Contiene metadatos, es decir, definiciones de otros objetos del sistema. Proporciona información acerca de:
 - ☐ Estructura lógica y física de la base de datos
 - ☐ Definición de todos los objetos de la base de datos.
 - ☐ Espacio asignado y utilizado por los objetos
 - ☐ Valores por omisión en las columnas.
 - ☐ Información acerca de las restricciones de integridad
 - ☐ Privilegios y roles otorgados a los usuarios
 - ☐ Estadísticas de utilización, frecuencias de consultas, de transacciones y el numero de accesos a diferentes porciones de la base de datos.



Sistema Manejador de Bases de Datos

Se trata de un software que facilita el proceso de **definir, construir, manipular y compartir** datos en una BD para diversas aplicaciones:

- **Definir** involucra especificar los tipos, estructuras y restricciones de los datos almacenados en la BD.
- **Construir** es el proceso de almacenar los datos en algún medio de almacenamiento controlado por el SABD.
- **Manipular** incluye funciones tales como consultas, recuperación y actualización de datos.
- **Compartir** es permitir el acceso a más de un usuario.

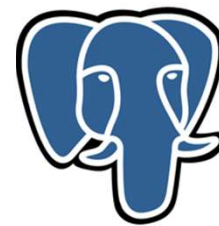


Sistema Manejador de Bases de Datos

ORACLE®

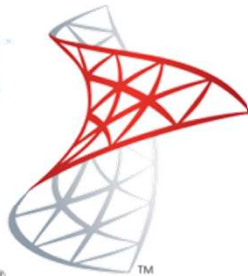


MariaDB



PostgreSQL

Informix



TERADATA®

Microsoft®
SQL Server®



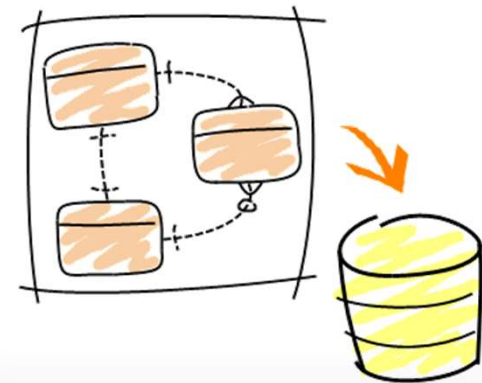


Sistema de Bases de Datos

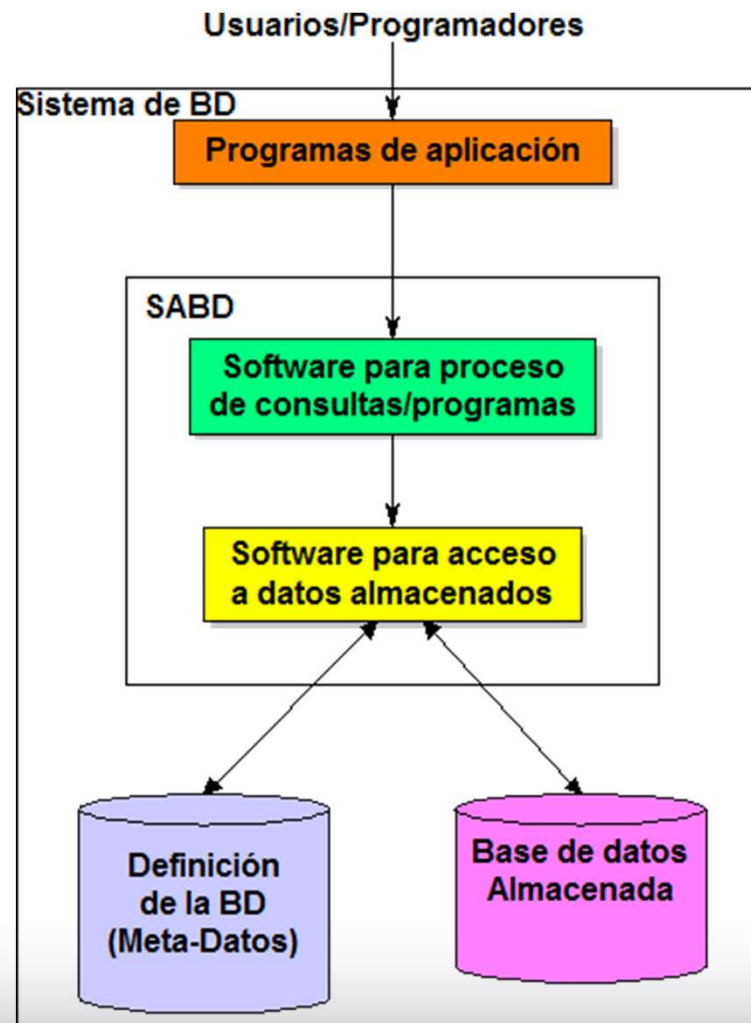
- Consiste en una colección de **datos interrelacionados** y un **conjunto de programas** para acceder a dichos datos.
- Su objetivo principal es proporcionar una forma de almacenar y recuperar la información de una base de datos de manera **práctica** y **eficiente**.

Sistema de Base de Datos = BD + SMBD

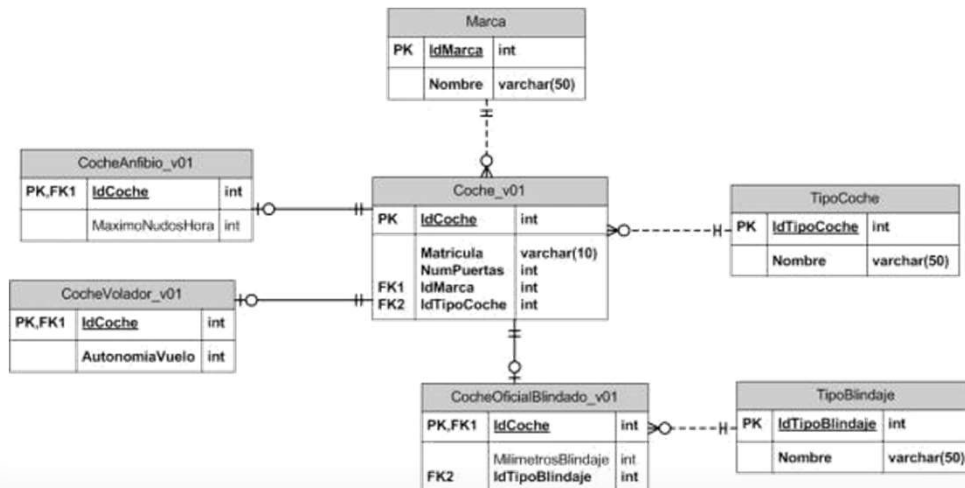
- Implica:
 - ☐ Definir estructuras de almacenamiento
 - ☐ Debe garantizar la fiabilidad (caídas o accesos no autorizados).
 - ☐ Proporcionar mecanismos de manipulación de la información.
 - ☐ Evitar datos anómalos



...Sistema de Bases de Datos

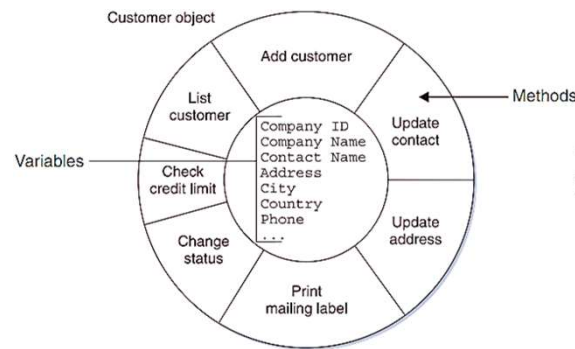
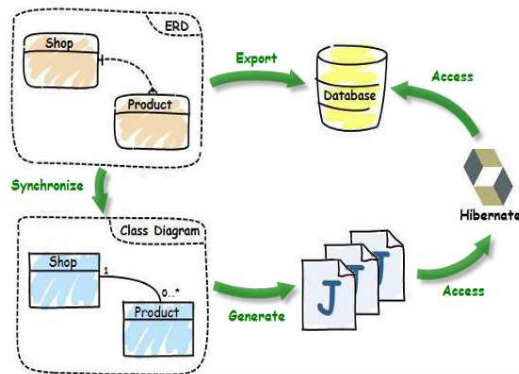


- **Modelo de datos.** Es un conjunto de conceptos que pueden usarse para describir la estructura de una BD. Proporciona tres tipos de herramientas:
 - ❑ Estructuras de datos para construir la BD.
 - ❑ Restricciones de integridad
 - ❑ Operaciones para trabajar con los datos.
- **Modelo relacional (70s).** Los datos se perciben como tablas y sólo hay tablas. Es un sistema cerrado en el sentido de que el resultado de las operaciones son siempre tablas.



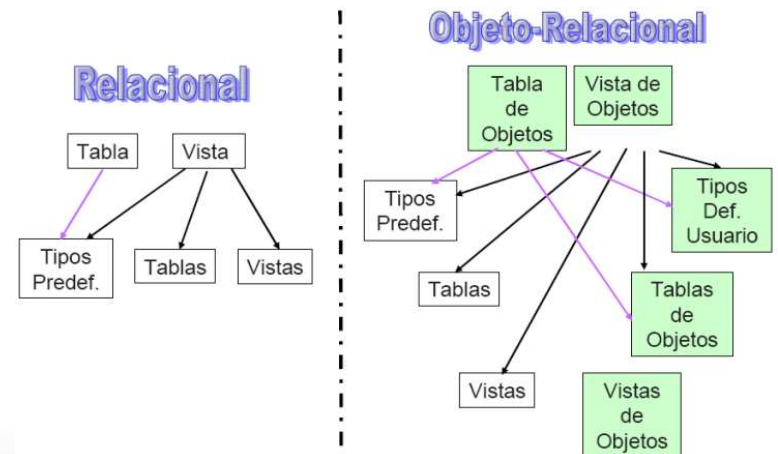
$$\text{SMBDR} = \text{SMBD} + \text{MR}$$

- **Modelo de objetos.** Los datos se modelan como objetos en los cuales, además del estado se tiene modelado su comportamiento.

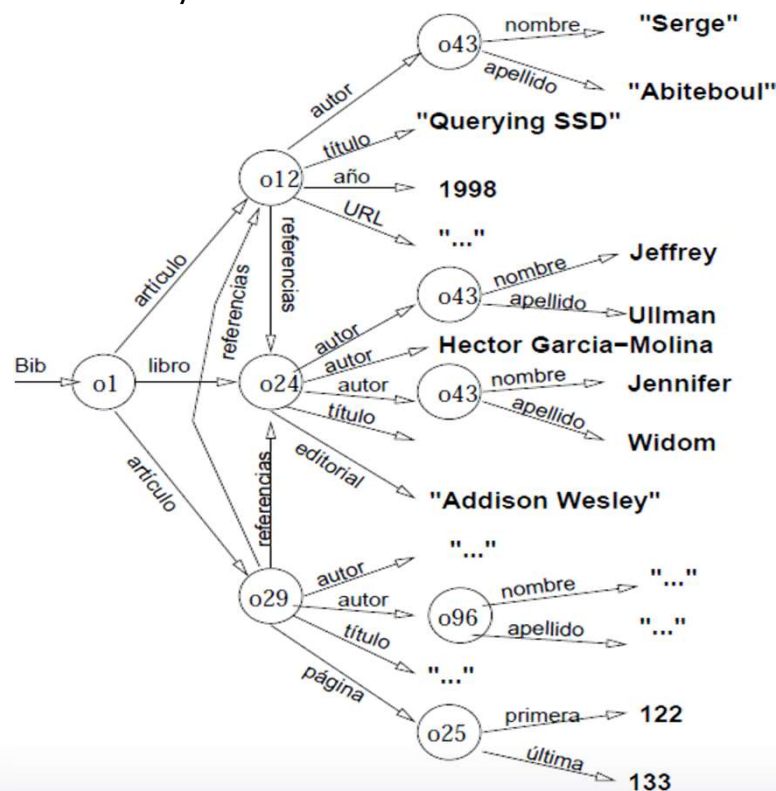


$$\text{SMBDO} = \text{SMBD} + \text{MO}$$

- **Modelo Objeto – Relacional (Oracle, SQL-99).**
Amplia las prestaciones de los RDBMS, adoptando ciertas características de la OO, manteniendo su funcionalidad, fiabilidad y escalabilidad



- **Modelo de datos semiestructurado.** Los datos no tienen una estructura rígida y mucho menos predefinida. Es una colección de nodos, donde cada nodo tiene datos con diferentes esquemas (bases de datos XML).



- Para trabajar con nuestras **BD**, los **SMBD** necesitan conocer su estructura: *qué entidades, qué atributos tendrán, qué tipos de datos habrá, etc.*
- El esquema de la BD es un elemento fundamental para la arquitectura, ya que permite independizar el **SMBD** \rightleftharpoons **BD**.
- En el periodo **1975-1982** el Comité **ANSI** (*American National Standards Institute*)/**SPARC** (*Standards Planning and Requirements Committee*) recomendó que la arquitectura de los **SMBD** previese **tres niveles** de descripción de la **BD**, que permitieran:
 - ☐ Separación de las aplicaciones de los datos
 - ☐ Manejo de múltiples vistas por parte de los usuarios
 - ☐ Uso de un catálogo para almacenar el esquema de la base de datos
- De esta forma, el SMBD debe poder garantizar la **transferencia de los datos** desde el nivel físico al nivel externo (transformación de datos o mapeo)



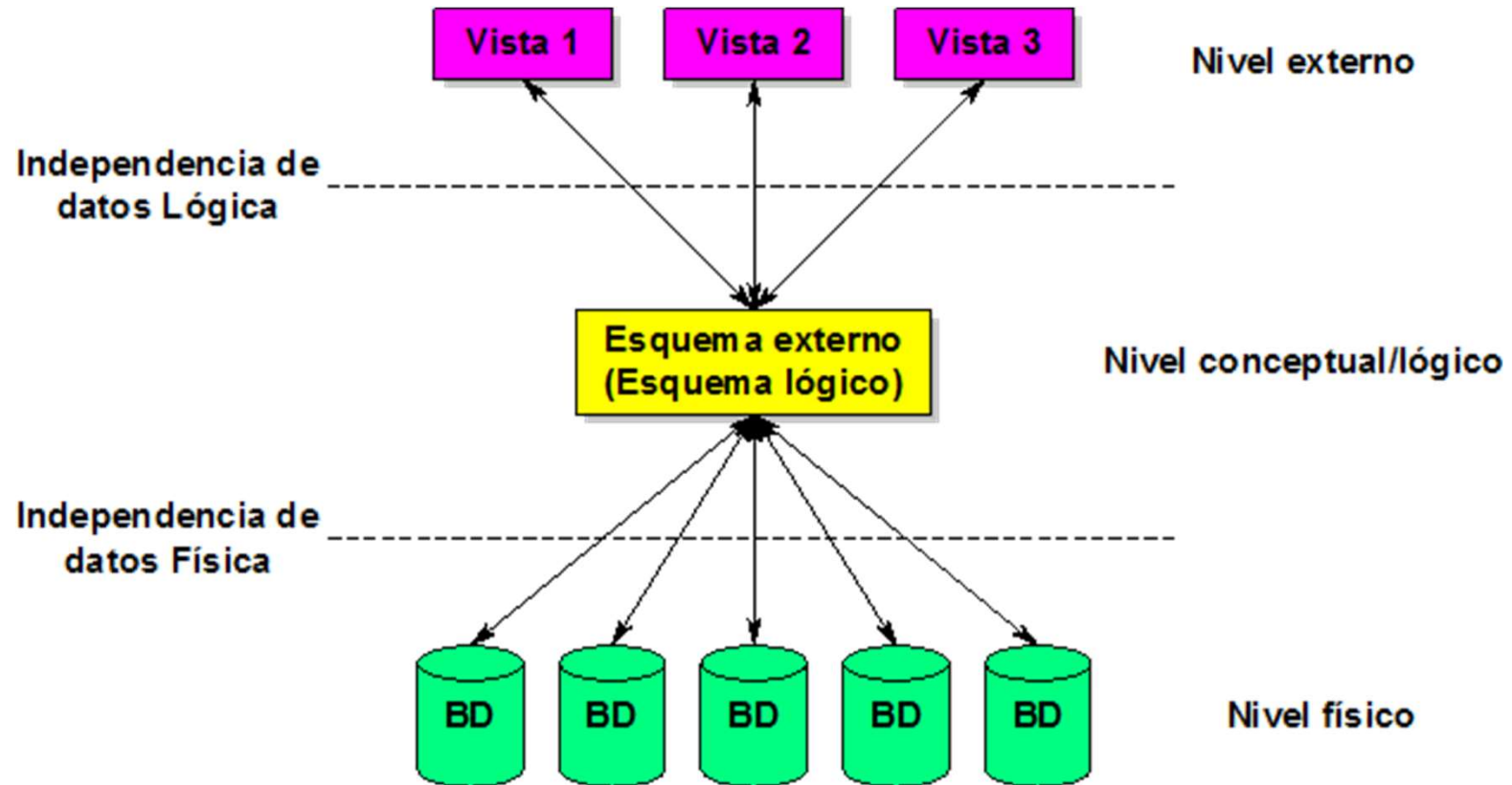
...Arquitectura ANSI/SPARK

- Los niveles que maneja esta arquitectura son:

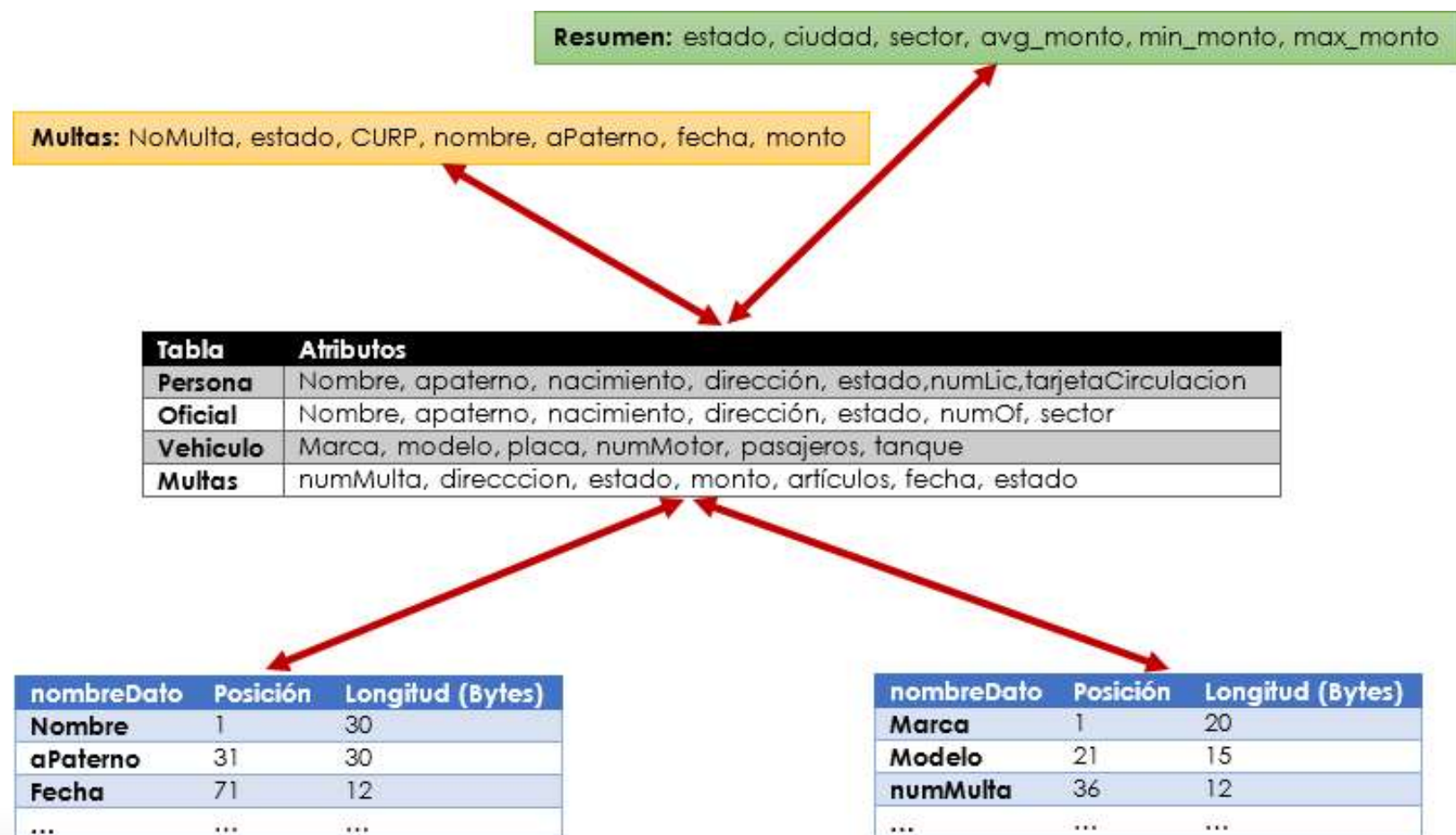
- ☐ **Nivel externo:** se sitúan las diferentes vistas lógicas que los procesos usuarios tendrán de las partes de la **BD** que utilizarán.
- ☐ **Nivel conceptual:** hay una sola descripción lógica básica, única y global, que sirve de referencia para el resto de los esquemas.
- ☐ **Nivel físico:** hay una sola descripción física, que denominamos esquema interno.

Nivel Lógico: oculta cómo se almacenan, mantienen y acceden

Nivel físico: qué índices, tamaño de página, etc.



...Arquitectura ANSI/SPARK

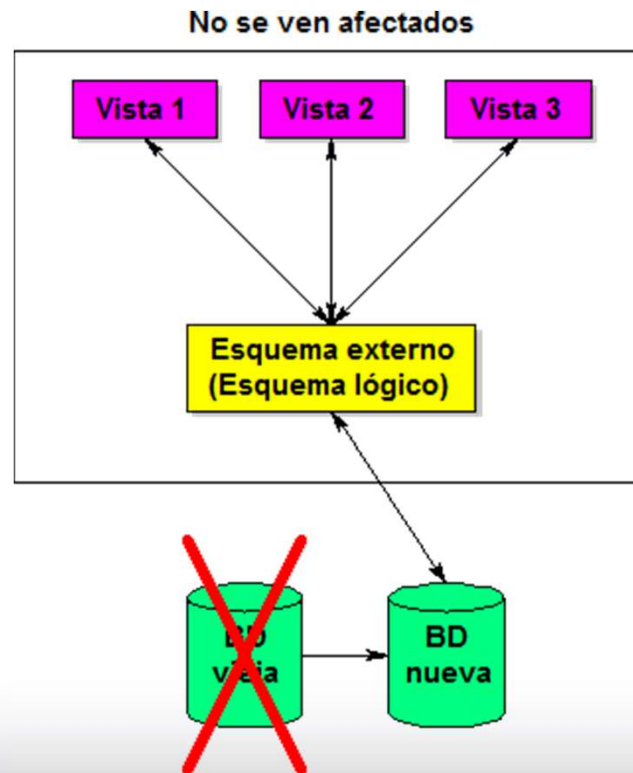


- Entre más niveles se tengan, se requiere ampliar el diccionario de datos, de manera que incluya información sobre cómo establecer correspondencia entre las peticiones de los usuarios y los datos en los diversos niveles.
- El SDBD debe poder garantizar la transferencia de los datos desde el nivel físico al nivel externo, a este proceso se le conoce como transformación de datos o mapeo:
 - Correspondencia conceptual/física
 - Correspondencia externa/conceptual



Independencia de datos. Capacidad de modificar una definición de esquema en un nivel, sin que esto afecte a las definiciones en esquemas en niveles superiores (o inferiores):

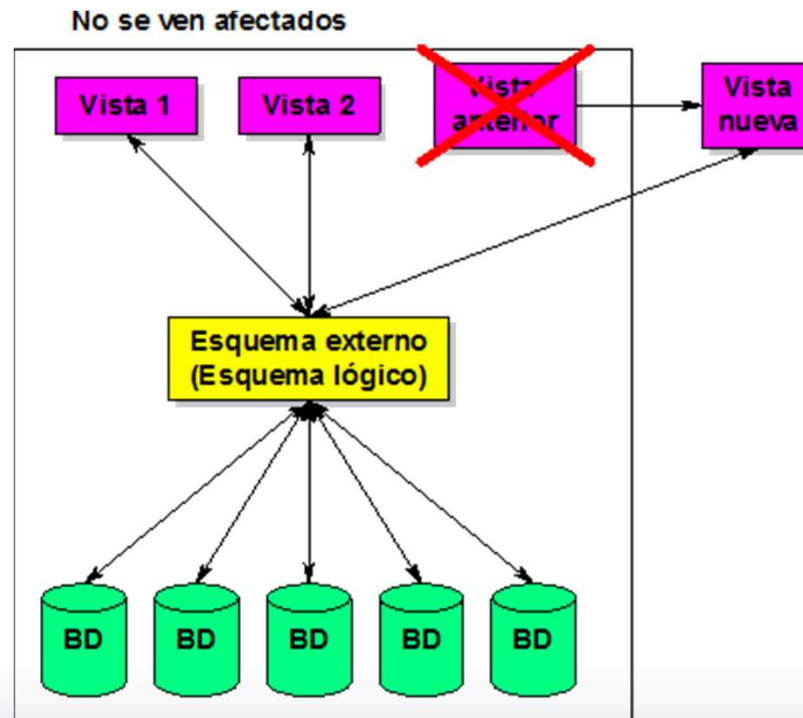
- **Independencia física.** Cuando los cambios en la organización física de la **BD** no afectan al mundo exterior.



- ❑ Cambios de un soporte de almacenamiento a otro.
- ❑ Cambio a otro “lugar” dentro del mismo soporte
- ❑ Cambio en el método de acceso a registros determinados, formato o codificación.

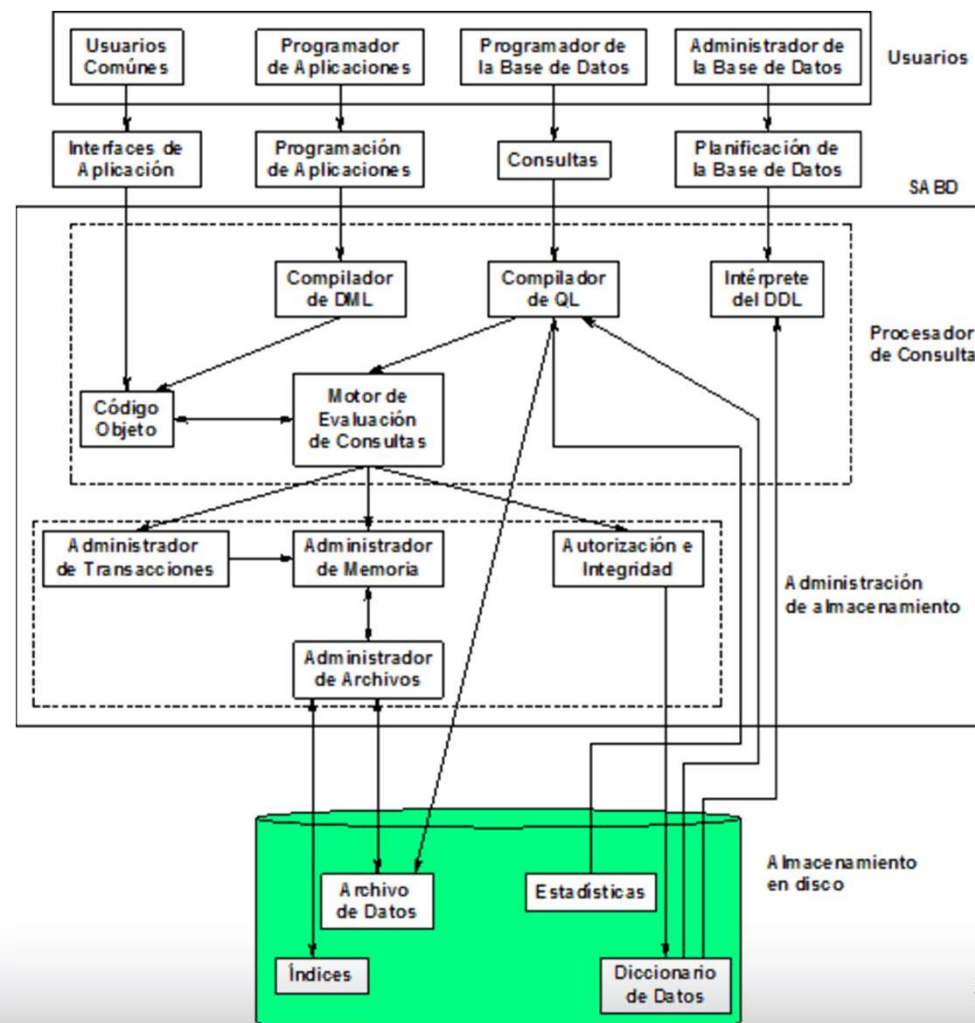
Independencia lógica. cuando los usuarios no se ven afectados por los cambios en el nivel lógico:

- ❑ **Cambios en el esquema conceptual:** No se afectan esquemas externo que no hagan referencia a entidades o atributos modificados.
- ❑ **Cambios en los esquemas externos**



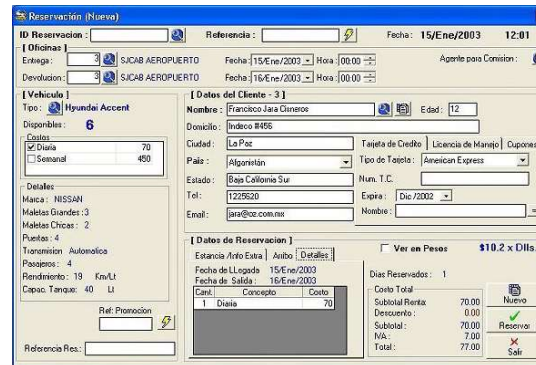
Funcionalidad de un SABD

- Lenguaje para definición de datos DDL
- Lenguaje para manipulación de datos DML
- Lenguaje para consulta de datos QL
- Administración de almacenamiento
- Administración de consultas
- Administración de transacciones
 - ☐ Control de concurrencia
 - ☐ Recuperación

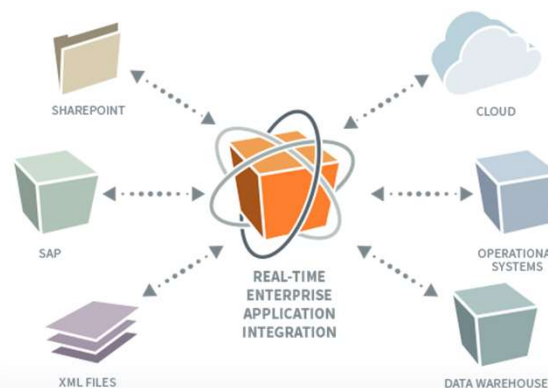


Aplicaciones tradicionales

- Sistemas de nóminas
- Sistemas de inventarios
- Sistemas estudiantiles
- Sistemas de reservaciones
- Sistemas bancarios, etc.

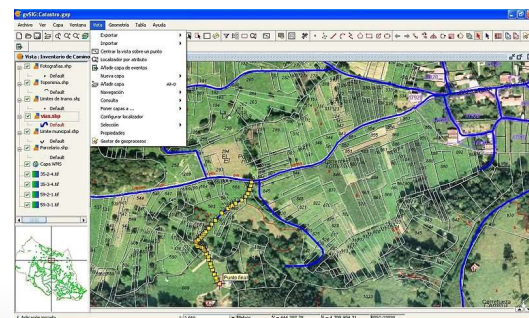
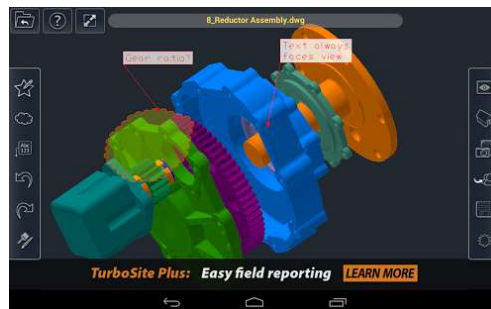
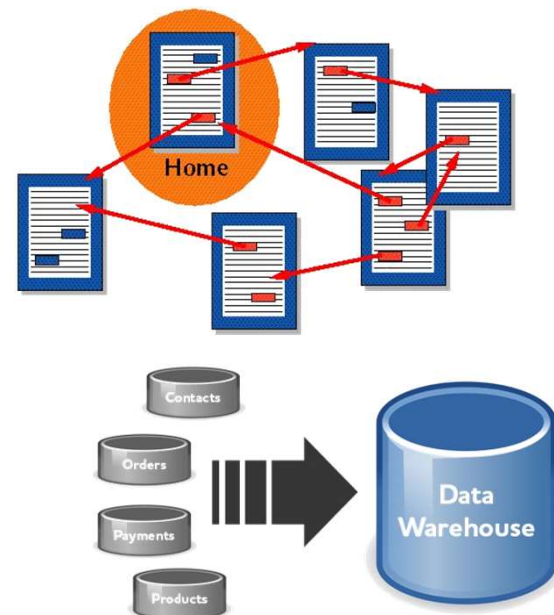


- Tienen necesidades y características que no pueden trabajarse con el modelo tradicional:
 - ❑ Otros tipos de datos (imágenes, sonido, texto, etc.)
 - ❑ Datos con estructura más compleja.
 - ❑ Grandes volúmenes de datos con interacciones más complejas
 - ❑ Conservar distintas versiones de un mismo diseño
 - ❑ Interoperación entre bases de datos preexistentes.



...Nuevas aplicaciones

- Aplicaciones CAD
- Sistemas multimedia
- Sistemas de información geográfica (GIS)
- Minería de datos
- Almacenes de datos (dwh)
- Aplicaciones de hipertexto, etc.



- **Modelado de requerimientos (conceptual, gráfico).**

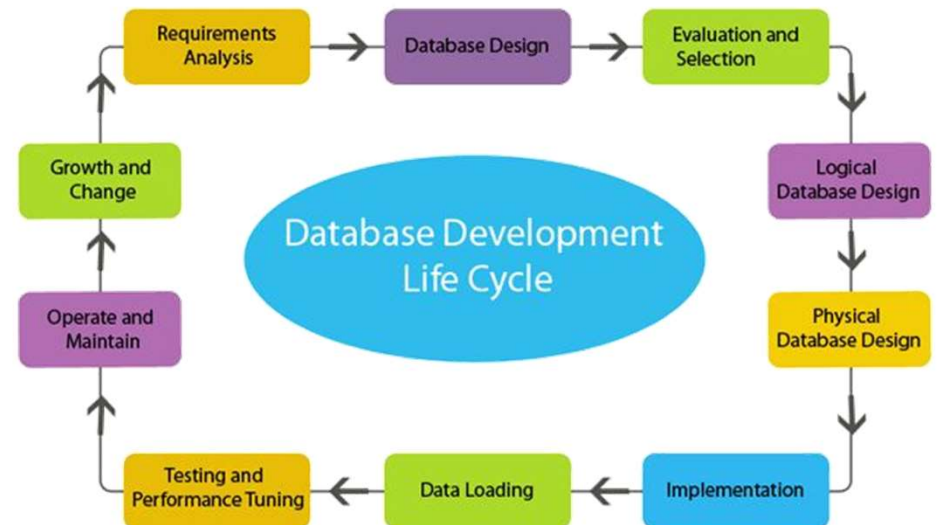
- ❑ Se decide cuáles entidades deben ser parte de la aplicación y cómo relacionarlas.

- **Diseño e implementación del esquema.**

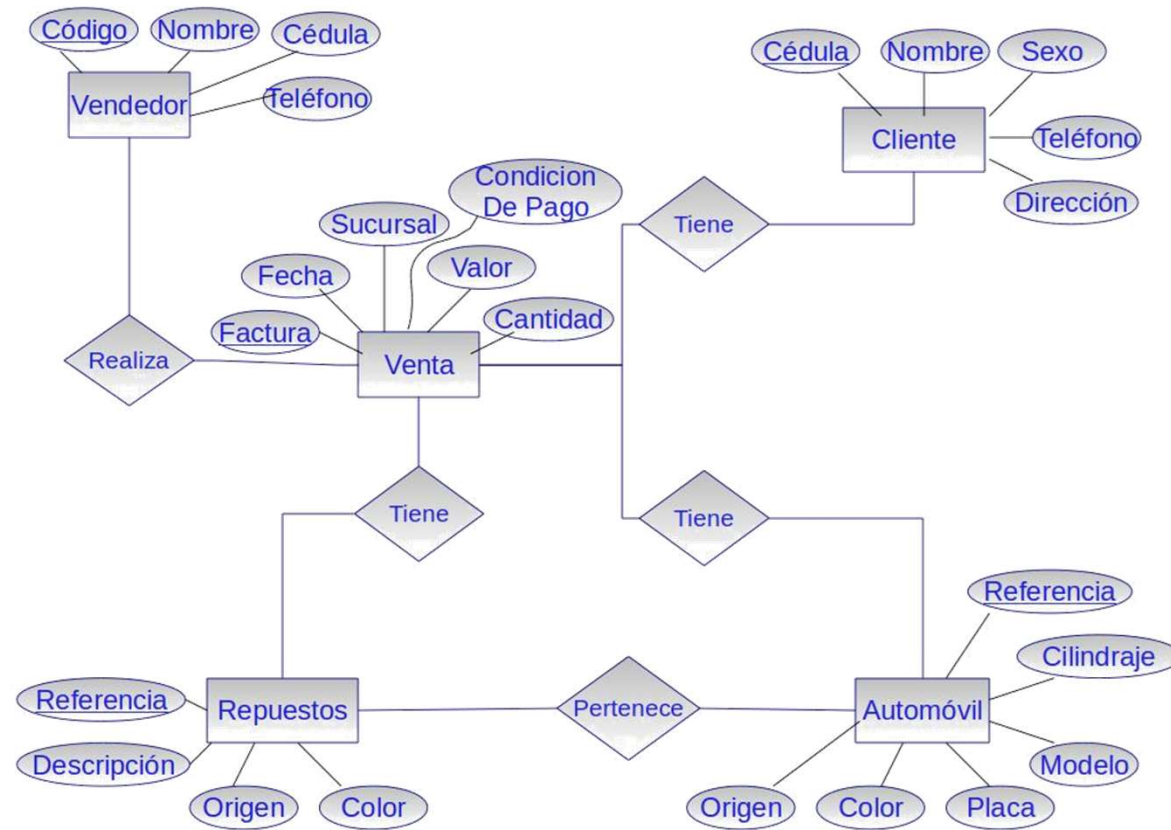
- ❑ Se decide qué repositorios se requieren.
 - ❑ Se implementan los repositorios en el SDBD.
 - ❑ Poblar los repositorios.

- **Escritura de la aplicación usando el SDBD**

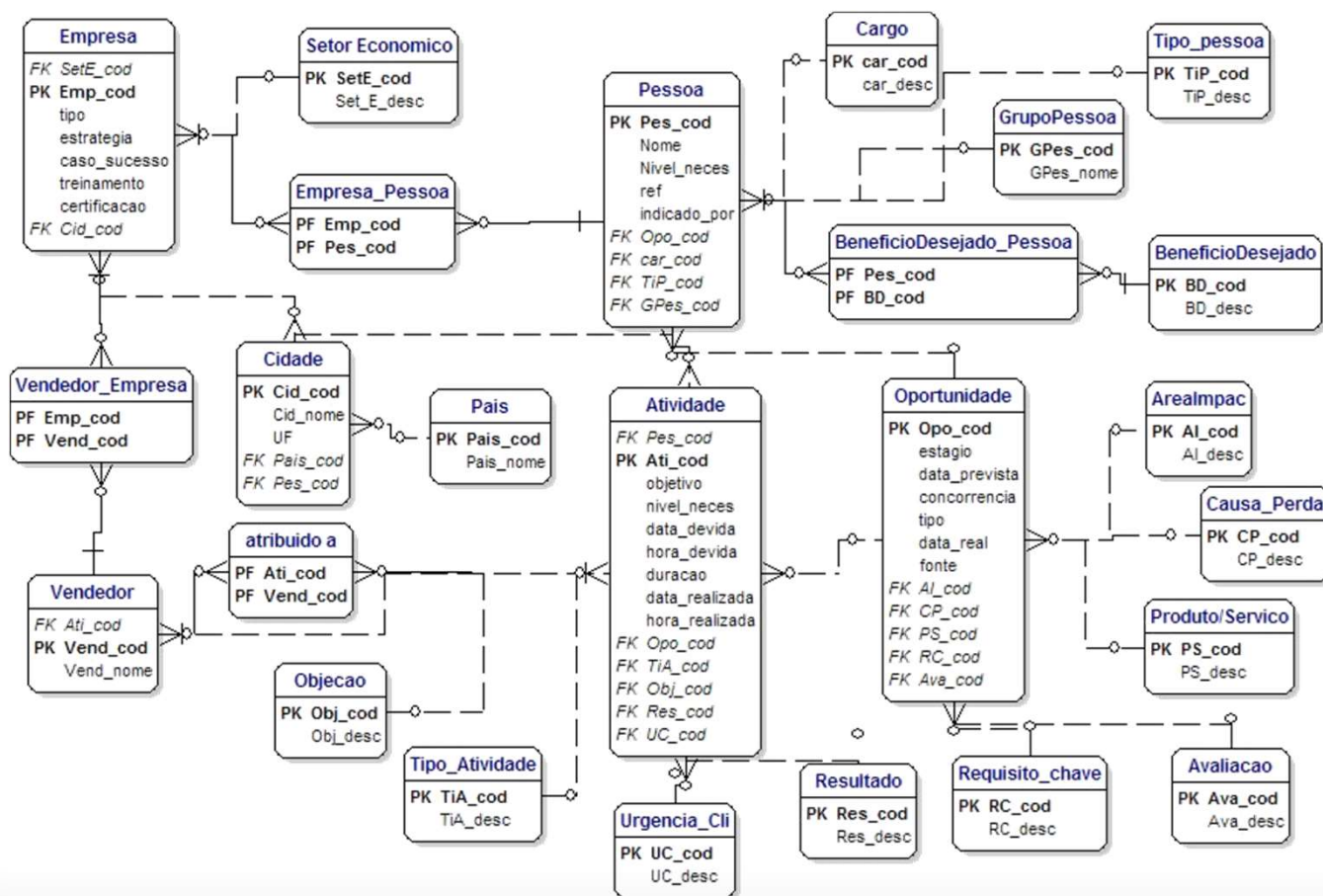
- ❑ Fácil pues ya no hay problemas con los datos.



Modelo conceptual



Implementación del esquema



Consultas a la base de datos

SQLQuery4.sql - (local)\...Cristian...)

```

select ArtNombre, ArtPrecio
From Articulo art Inner Join Categoria cat
on art.CatId = cat.CatId
Where CatNombre = 'Bebidas'

```

Resultados Mensajes

	ArtNombre	ArtPrecio
1	Cerveza negra Steeleye	18.00
2	Cerveza Sasquatch	15.00
3	Vino Côte de Blaye	250.25
4	Cerveza Heineken	20.50
5	CocaCola	5.75

SQLQuery2.sql - VM...administrator (64))

```

SELECT FORMAT(GETDATE(), 'dd/mm/yyyy') as Date

DECLARE @NUM NUMERIC(9,2) = 59720.34

SELECT
    FORMAT(@NUM, 'c', 'en-US') as US
    ,FORMAT(@NUM, 'c', 'es-ES') as SP
    ,FORMAT(@NUM, 'c', 'en-GB') as UK

```

Results Messages

	Date
1	08/18/2013

	US	SP	UK
1	\$59,720.34	59.720,34 €	£59,720.34

SQLQuery1.sql - VM...administrator (63))

```

SELECT CONCAT(Nombre, ' ', Ape1, ' ', Ape2) AS NombreCompleto
FROM
(
    SELECT 'Pedro' AS Nombre, 'Martín' AS Ape1, 'López' AS Ape2
    UNION ALL
    SELECT 'Ramón' AS Nombre, 'Pérez' AS Ape1, NULL AS Ape2
) AS Personas

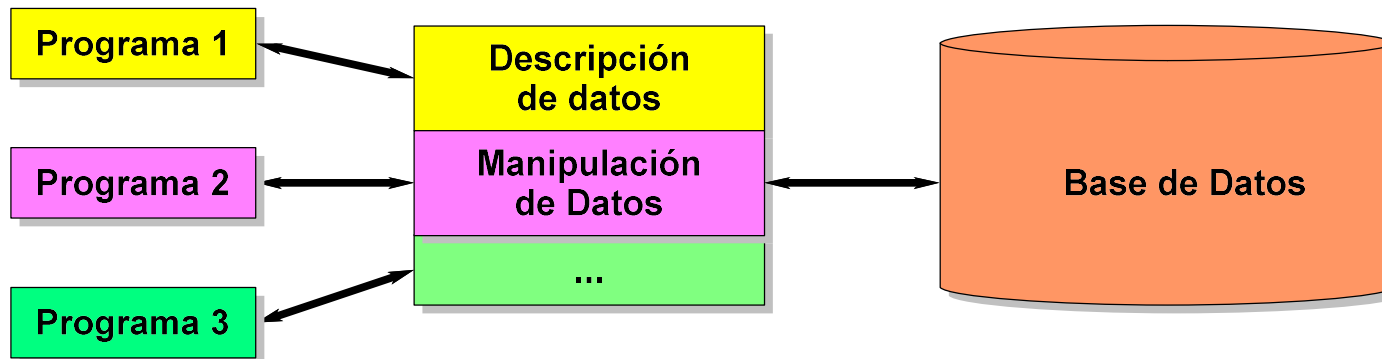
```

Results Messages

	NombreCompleto
1	Pedro Martín López
2	Ramón Pérez

Enfoque de las Bases de Datos

- Un solo depósito de datos definido una sola vez y puede accederse por varios usuarios.
- Representación de relaciones complejas entre datos; el SABD tiene la BD y su descripción completa.
- Los programas de acceso se escriben con independencia de archivos específicos.
- Control de redundancia, restricciones de acceso, restricciones de integridad.
- Proporciona respaldo y recuperación; reducción de tiempo de desarrollo; flexibilidad; disponibilidad de datos actualizados.



No estés muy orgulloso de haber comprendido estas notas.
La habilidad para manejar los conceptos básicos es insignificante comparado con el poder de la Fuerza.

