



FACULTAD DE CIENCIAS Y TECNOLOGÍA

Ingeniería en Ciencias de la Computación
Ingeniería en Tecnologías de la Información y
Seguridad

Capítulo I

SISTEMAS DE BASES DE DATOS

Ing. Edgar T. Espinoza R.



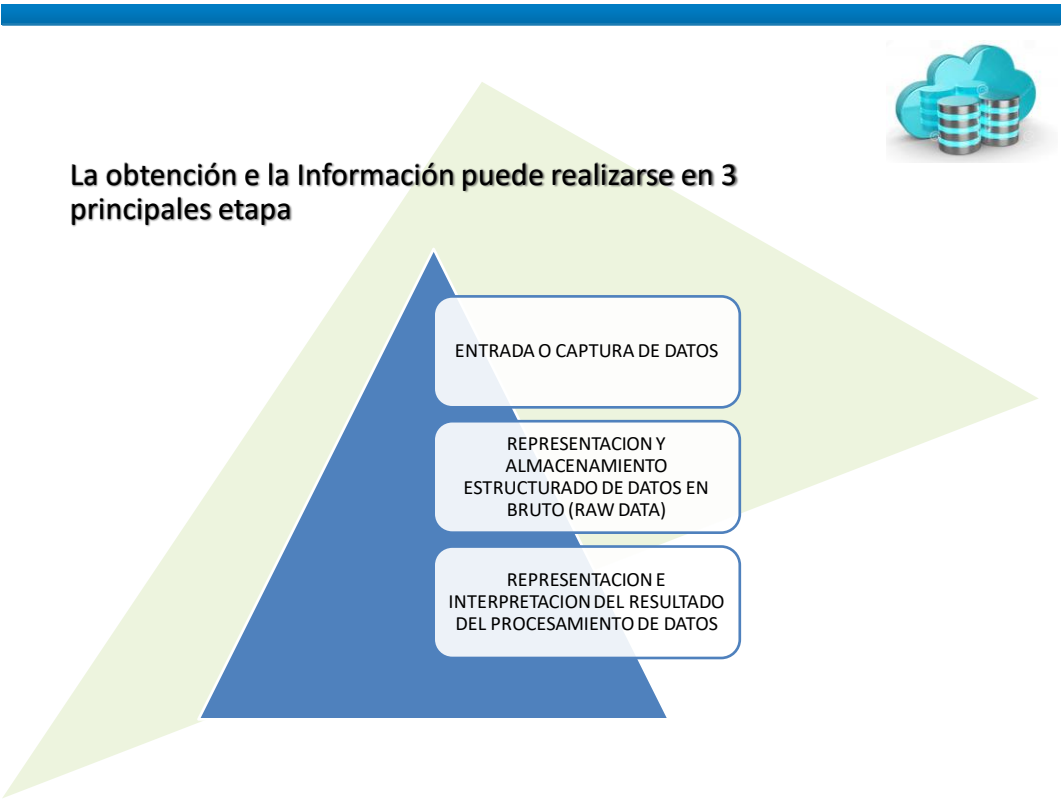
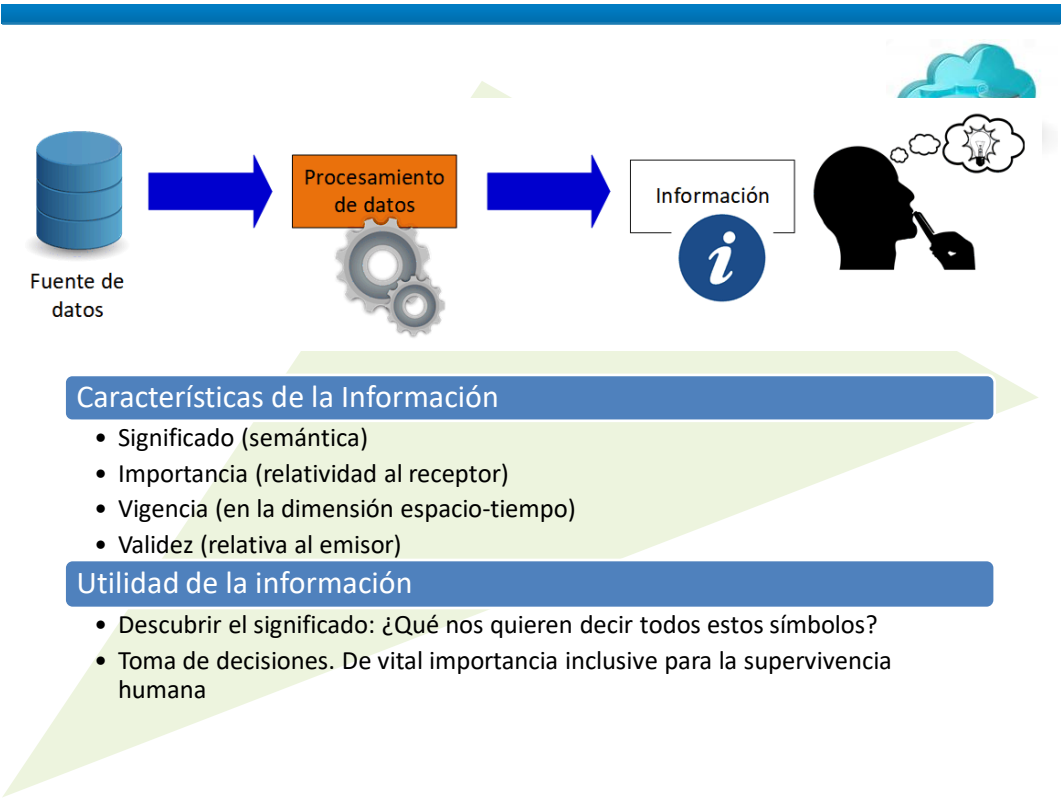
1. Datos e Información



¿Por qué los datos son tan importantes como para dedicar un semestre a estudiar la forma en la que estos se administran?

- Representan la *herramienta principal* para lograr el correcto funcionamiento y obtención de beneficios para cualquier organización, institución, compañía, etc.
- Imaginar a un negocio tratar de realizar sus operaciones sin conocer ¿quiénes son sus clientes?, ¿qué productos venden?, ¿quiénes son sus deudores? Toda esta información debe estar *disponible* para que el negocio pueda operar de forma correcta.
- La cantidad de datos puede ir desde unos cuantos MB de datos producidos al día hasta volúmenes inimaginables, Terabytes o Petabytes de datos nuevos generados en unas cuantas horas.





2. Administración de datos



Para obtener información útil y correcta surgen algunos requerimientos:

- Almacenamiento de datos en estructuras adecuadas, representados por un “tipo de dato” y formateados de forma homogénea. ¿Será posible procesar estos datos?
- Manejo de volúmenes diversos de datos.
- Uso compartido y concurrente.
- Aseguramiento de la permanencia de datos.
- Eficiencia al almacenar o al recuperar los datos.
- Ayuda a garantizar la integridad de los datos.
- Respallos
- Seguridad
- Interpretación de lenguajes de acceso a datos, interfaces de programación
- Interfaces de comunicación: Local, remota, etc.



Para realizar lo anterior, es necesario realizar una Administración de los datos

Se requiere contar con una herramienta sistematizada que sea capaz de proporcionar una solución viable a cada uno de los puntos antes mencionados. A esta herramienta se le conoce como *Sistema Administrador de Bases de Datos: DBMS*.

3. Base de Datos:



- Colección de datos interrelacionados que representan información de interés para un sistema de información y/o usuario final.
- Los datos que constituyen a la base, se almacenan de manera sistemática para su posterior uso a través diversas estructuras de datos.
- Símbolos comúnmente empleados para representar bases de datos:



La colección e datos almacenados es administrada por un DBMS, y se divide en 2 grupos:



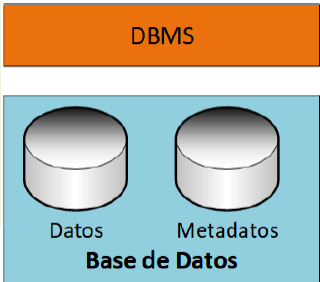
Datos de interés para el usuario (datos del usuario)

▪Representan la fuente para la generación de información de interés para el usuario final.

Metadatos (Datos acerca de los datos).

•Proporciona una descripción de las características de los datos almacenados: o Tipos de datos.

- Describen las relaciones existentes entre los datos.
- Restricciones de los datos, etc.
- Los metadatos representan el llamado **Diccionario de datos**.



4. Database Management System (DBMS)



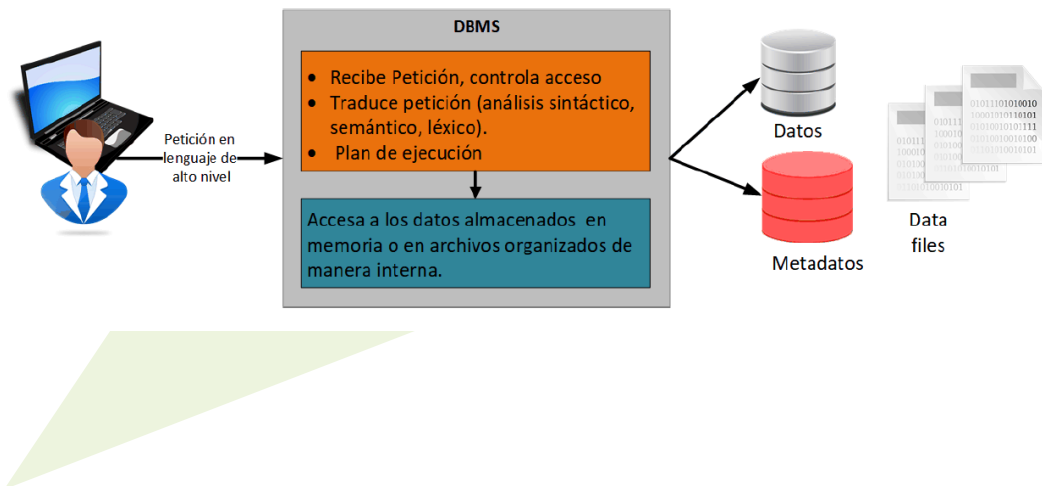
Colección programas (software) encargado de administrar la estructura de la base de datos y del control de acceso a los datos almacenados

Principales funciones:

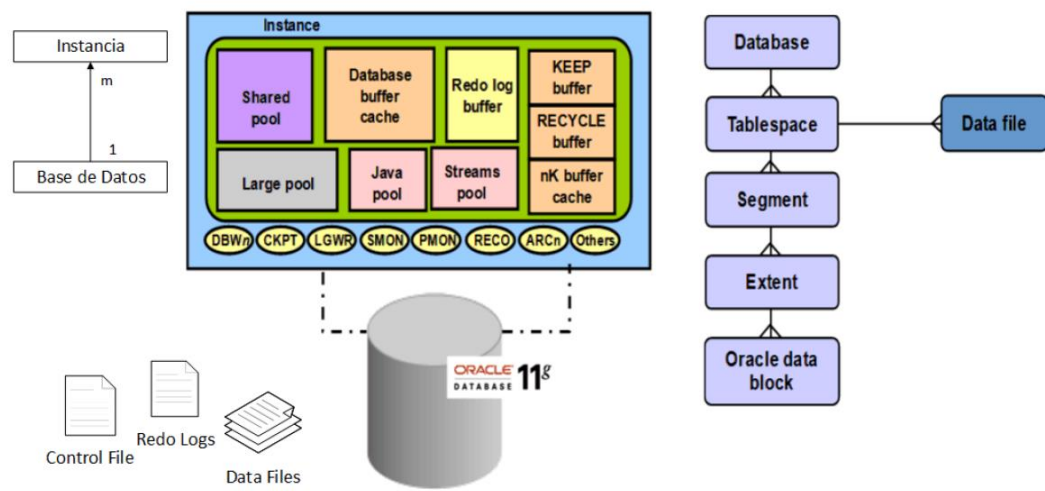
- Actúa como intermediario entre el usuario final y la base de datos.
- Implementa los requerimientos antes mencionados.
- La estructura del BD es almacenada en un conjunto de archivos, típicamente llamados **Data files**.
- Esta estructura de dichos archivos es *independiente* y desconocida por el usuario. La única forma para acceder a los datos es usando un DBMS.
- Se encarga de realizar la administración del diccionario de datos.



Arquitectura básica de un DBMS



Ejemplo, Arquitectura Oracle 11g



5. Tipos de Bases de Datos

En general, las clasificaciones más comunes que suelen aplicarse son las siguientes:

- **Por el número de usuarios, las bases de datos se clasifican en:**
 - De usuario único (Single User Database).
 - Base de datos para un grupo de trabajo (Workgroup database), aprox. hasta 50 usuarios
 - Base de datos empresarial (Enterprise Database), más de 50 usuarios.
- **Por la ubicación:**
 - Base de datos centralizada.
 - Base de datos distribuida.
 - Base de datos en la nube.
- **Por el uso:**
 - Base de datos operacional, transaccional o de producción OLTP (OnLine Transaction Processing).
 - Data warehouse OLAP (On-Line Analytical Processing).

▪ Por el modelo de datos:

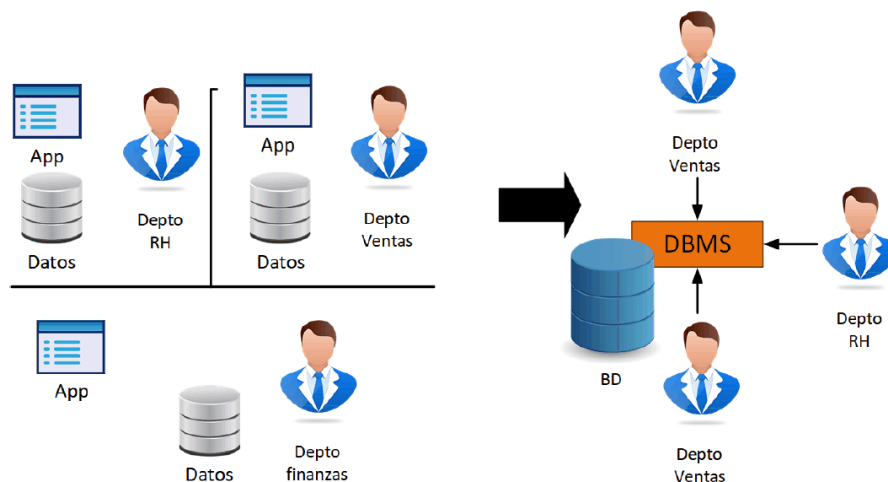
- Bases de datos Jerárquicas.
- Bases de datos de Red.
- **Bases de datos Relacionales** (La que se estudiará en el curso): **RDBMS**
 - Bases de datos relacionales que incorporan conceptos de la programación orientada a objetos (**ORDBMS**)
 - Bases de datos relacionales para manejo de documentos XML
 - Bases de datos relacionales para manejo de documentos JSON
 - Bases de datos relacionales para el manejo de grafos
- Bases de datos Orientadas a Objetos (**OODBMS**).
- Bases de datos multidimensionales.
- Bases de datos NoSQL

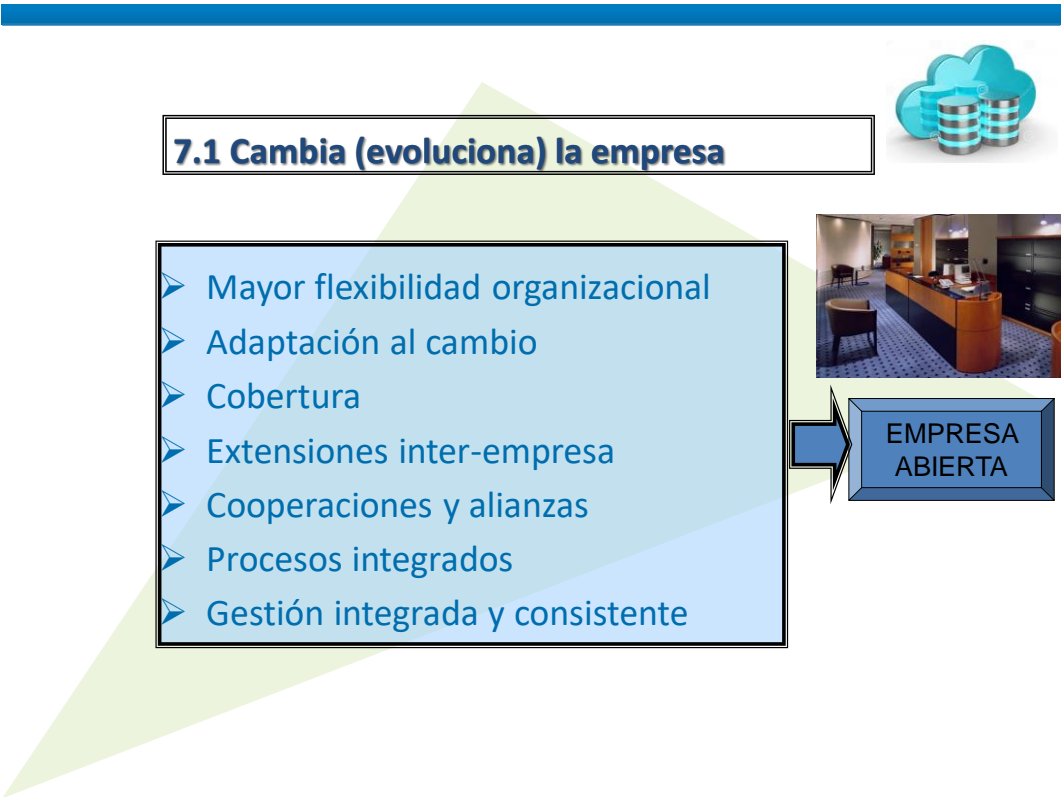
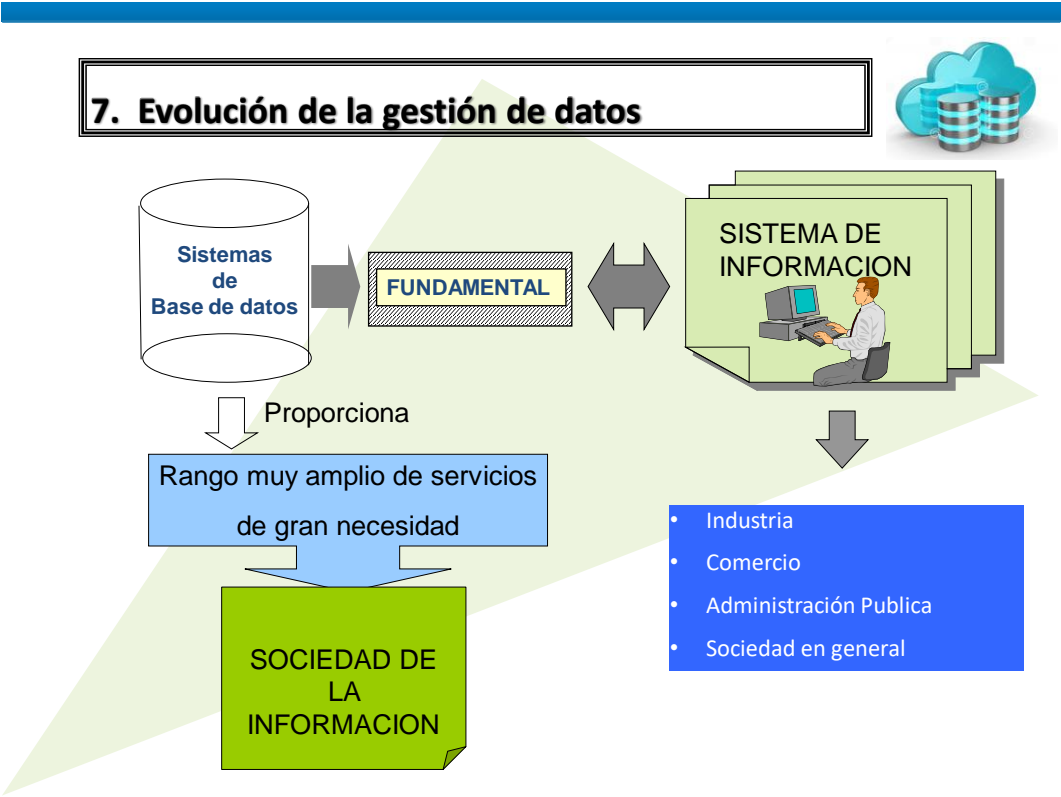


6 . Sistemas de Base de Datos



Diferencia entre el uso de un sistema de archivos, y el uso de un sistema de base de datos en una empresa formada por diferentes departamentos. Cada uno de ellos realiza su propia administración de datos.





7. 2. Hardware Y Comunicaciones



Leyes de evolución del Hardware

- Ley de Moore
- Ley de Hoagland
- Ley de Joy
- Predicción de Gray

Hardware cada vez más barato y más potente

Maquinas paralelas

Procesamiento paralelo

Técnicas de compresión

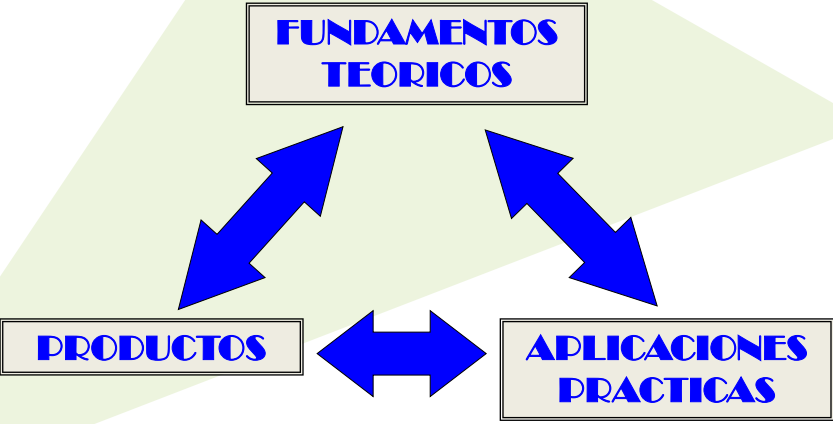
Dispositivos de almacenamiento ópticos

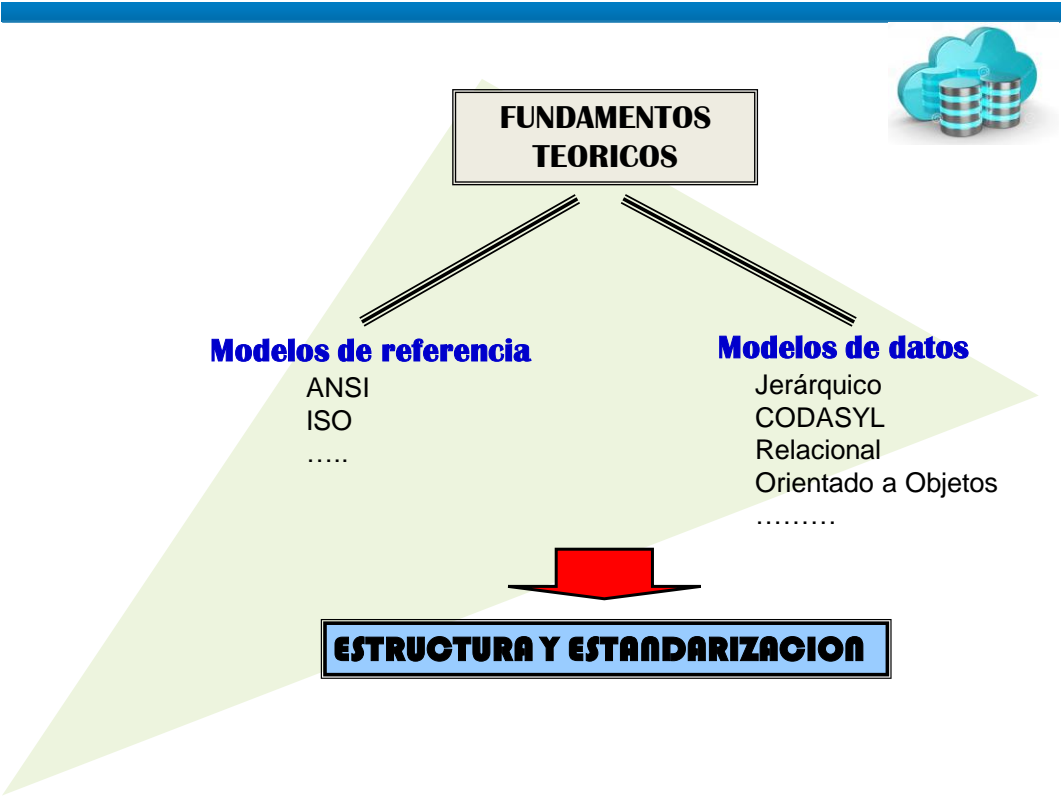
Digitalizadores de Audio, Video, etc.

8 . Líneas de evolución



8.1. Factores claves

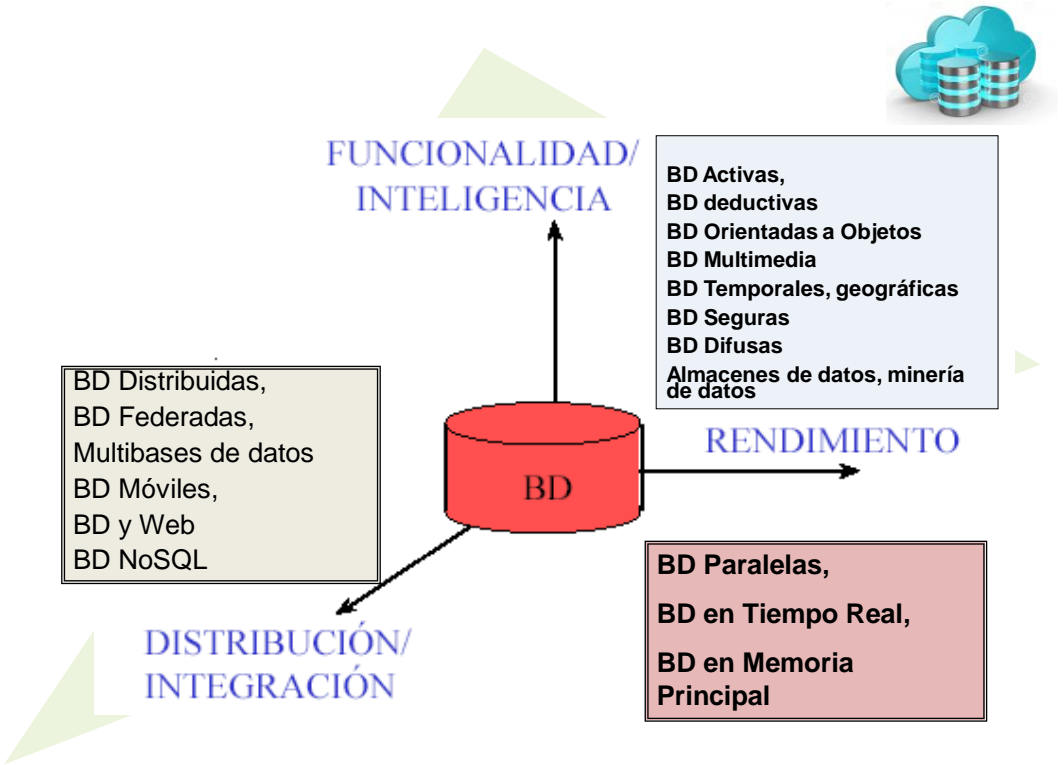


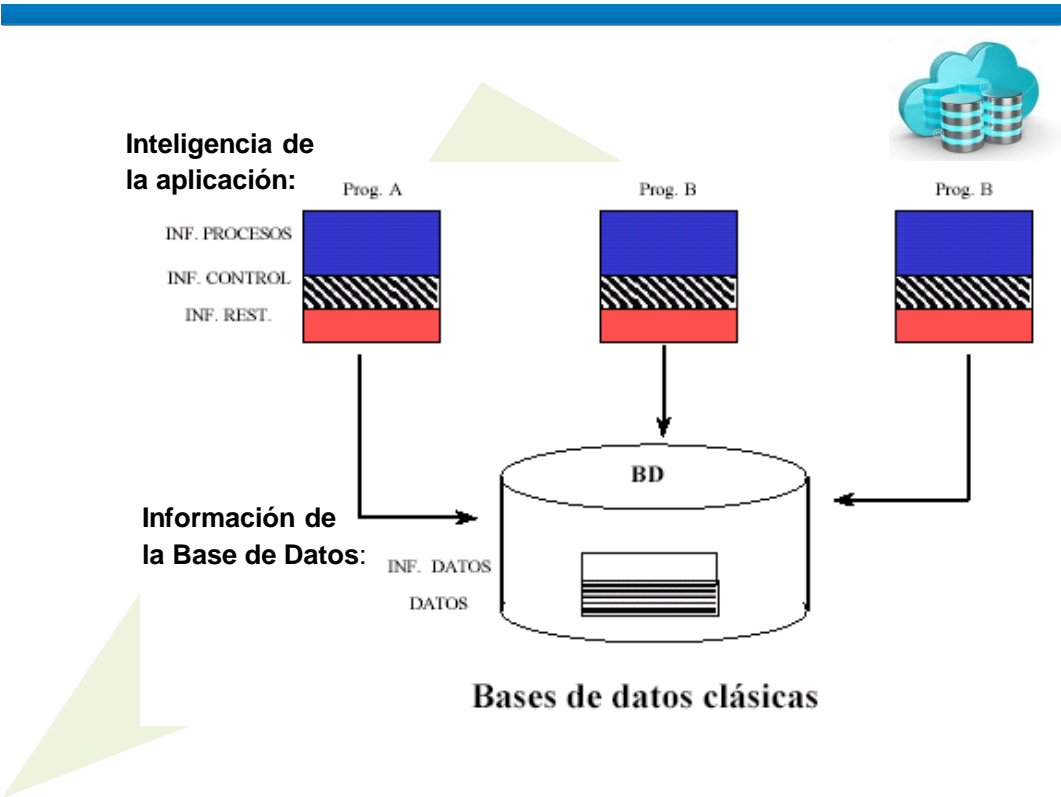
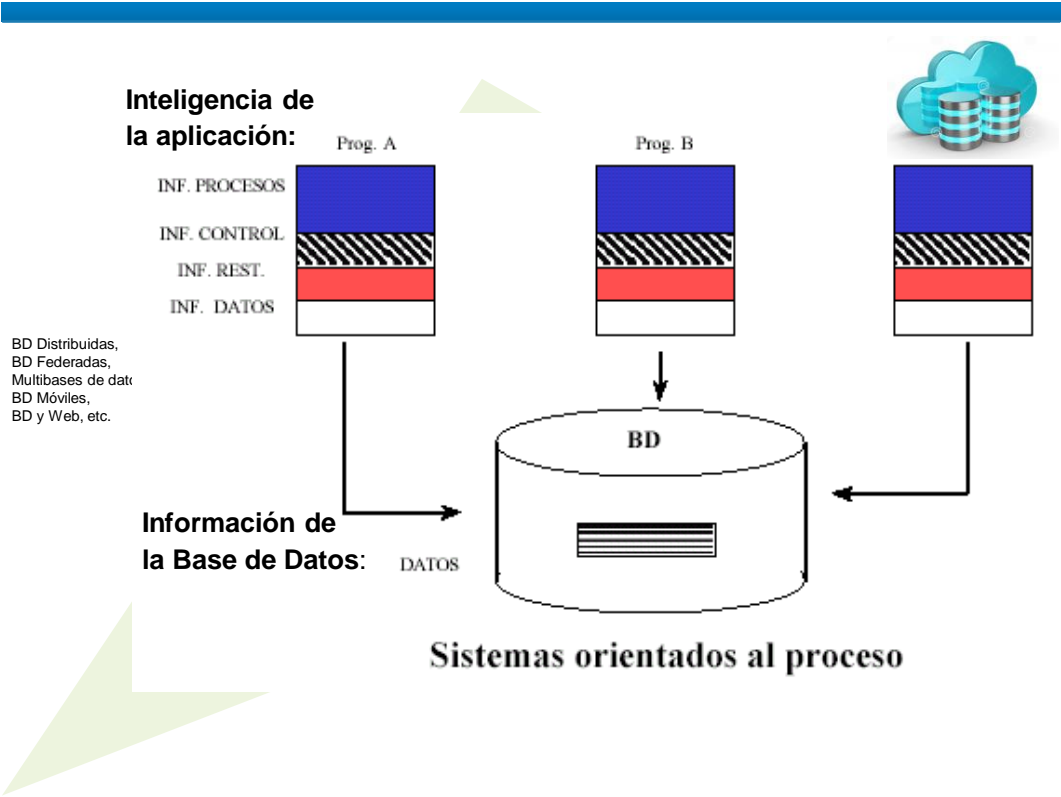


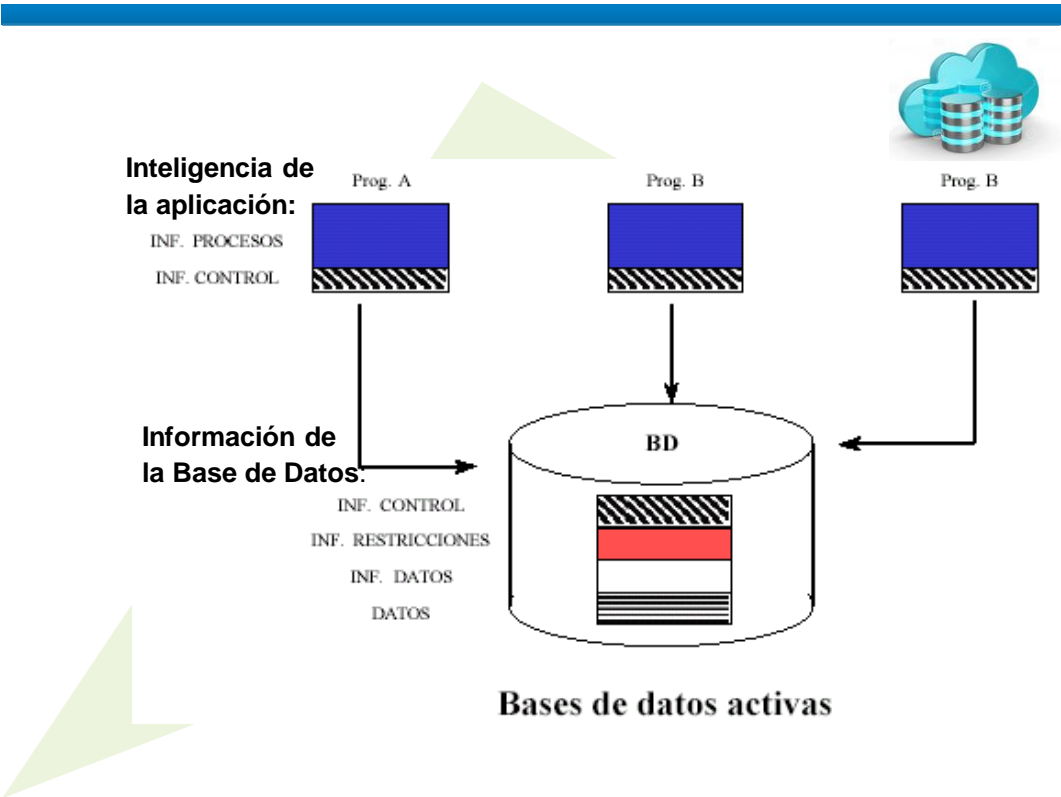
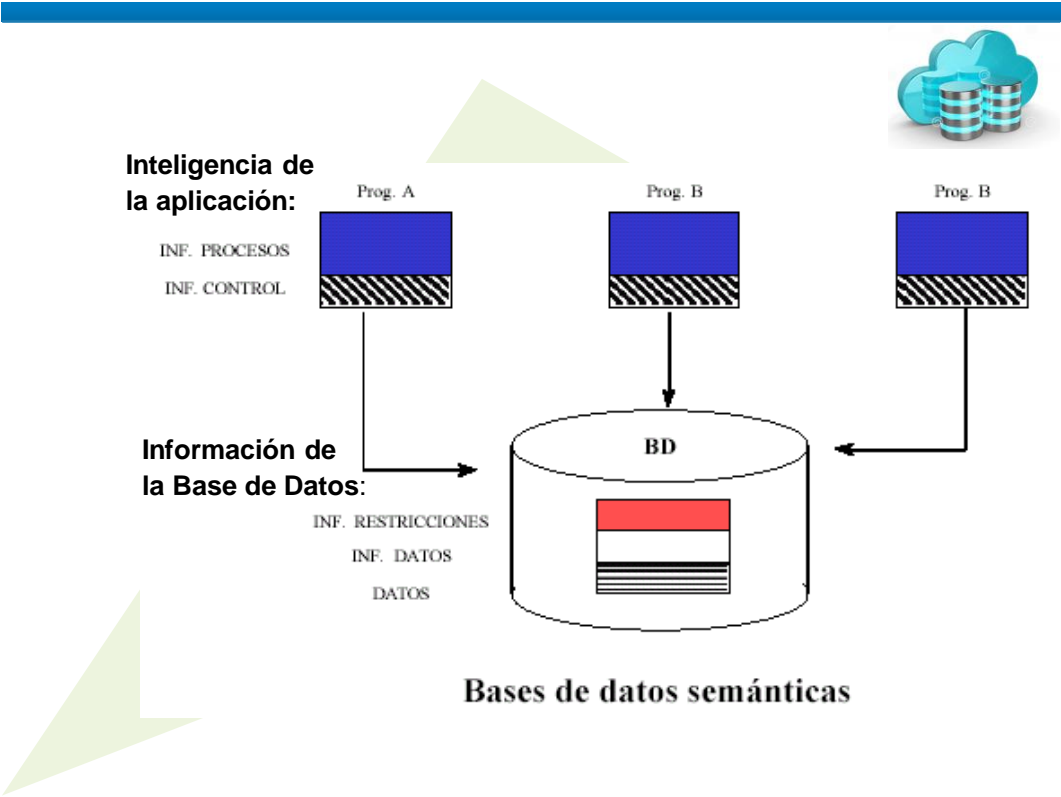


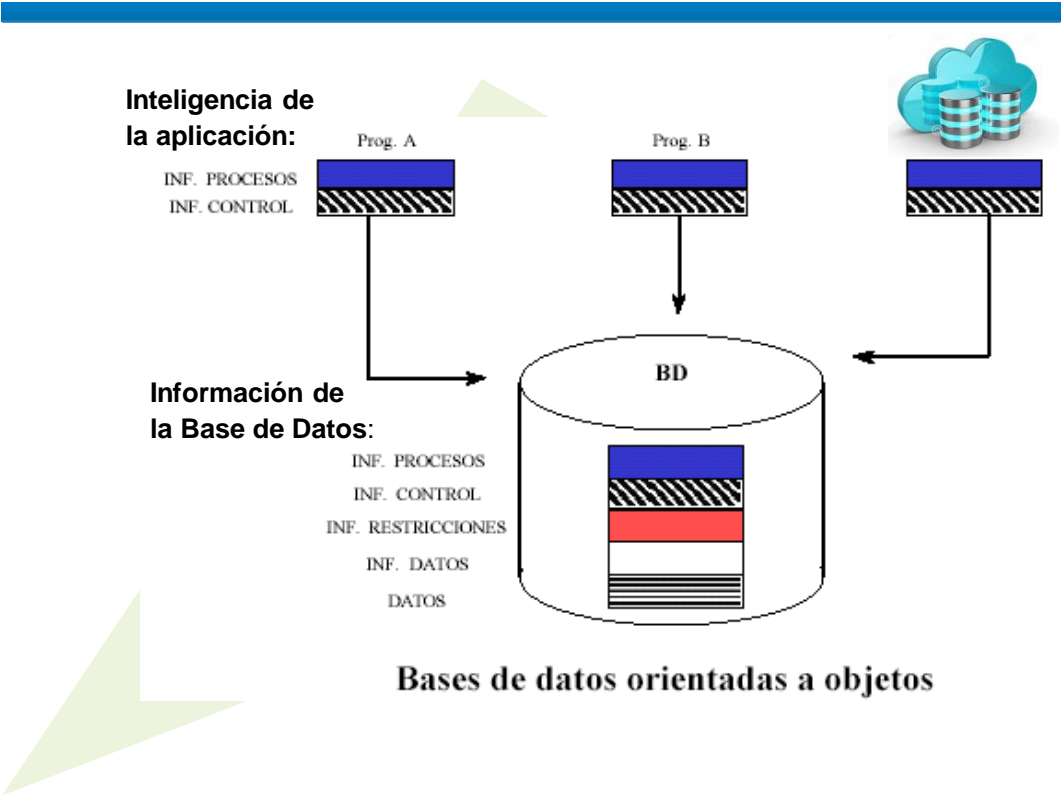
8.2.Dimensiones:

A la hora de clasificar los avances en el campo de las Base de Datos, podemos identificar tres dimensiones









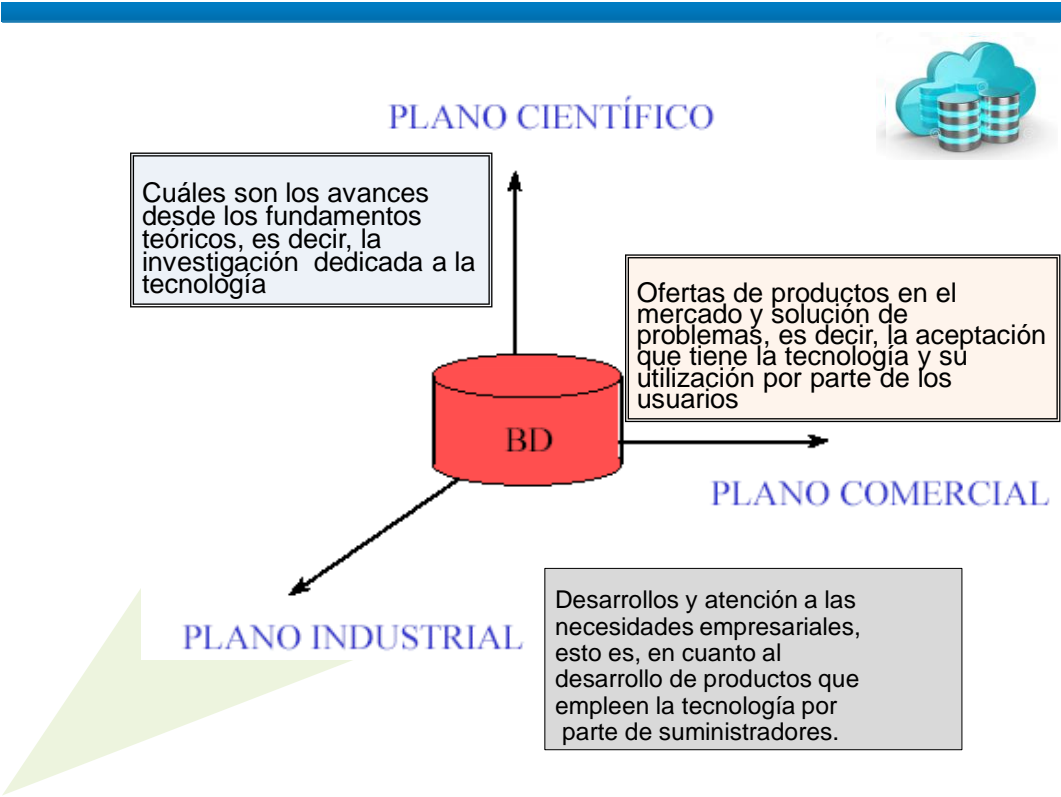
9. Madurez de las nuevas tecnologías



Niels Bohr afirmaba: *“Predecir es muy difícil y sobre todo el futuro”*.

En los primeros próximos diez años, sería razonable que maduraran algunas de las tecnologías citadas y que todavía hoy se encuentran poco desarrolladas.

Esta madurez es analizada y producida en los tres planos siguientes (De Miguel y Piattini 1999):





10 . Elementos de un Sistemas de Base de Datos




- **Hardware** (servidores, servidores de almacenamiento, componentes de red).
- **Software**
 - Sistemas operativos
 - DBMS
 - Programas de utilería, administración, monitoreo, etc.
- **Personal**
 - Administradores del sistema operativo.
 - Usuario final.
 - Rol asociado con la base de datos (ver siguiente tabla).


11. Roles asociados con Bases de Datos





Nombre del rol	Descripción	Principales habilidades.
Desarrollador / Programador de bases de datos.	Crea y mantiene aplicaciones desarrolladas con herramientas, extensiones y lenguajes de programación de bases de datos.	Programación SQL, extensiones PL/SQL, etc.
 Analista de bases de datos.	Generación de los artefactos necesarios para extraer y comprender los requerimientos tanto funcionales como no funcionales de un usuario o cliente final.	Facilidad de interacción con el usuario final, capacidad de conciliar y proponer estrategias y alternativas de solución al usuario final. En algunos casos representa el medio de comunicación entre el equipo de desarrollo y el usuario final.
 Diseñador de la base de datos.	A partir de un caso de estudio o problema real, generar diferentes modelos de datos que describen la estructura y el comportamiento de los datos tratando de conservar un equilibrio entre desempeño, niveles de normalización, etc.	Diseño de modelos de datos, conceptos sólidos de programación orientada a objetos, Ingeniería de software, SQL.



Nombre del rol	Descripción	Principales habilidades.
 Administrador de la base de datos (DBA)	Encargado de mantener la buena salud del DBMS para que opere de forma más óptima posible Tarea: Leer Artículo Responsabilidades del DBA en tiempos 2.0. Obtener el artículo de: <ul style="list-style-type: none">http://dbagroup.cl/blog/?p=200	Conocimientos sólidos en arquitecturas y algoritmos internos de la base de datos, estructuras físicas y lógicas tanto de almacenamiento como de memoria y procesos, además de contar con fundamentos sólidos en áreas como: Sistemas operativos, redes, diseño y programación de sistemas de software, seguridad, etc.
Especialista en análisis de datos.	Sus conocimientos le permiten aplicar tecnologías, técnicas, cálculos matemáticos y estadísticos, así como el uso de una diversidad de herramientas enfocadas al análisis de datos de grandes volúmenes. Data warehouse , Big Data, Ciencia de datos.	Conocimientos sólidos en las áreas de estadística, inteligencia artificial, así como el uso de diversas herramientas y tecnologías enfocadas al análisis de datos que permitan descubrir conocimiento y con ello la toma de decisiones importantes para una empresa u organización.



Nombre del rol	Descripción	Principales habilidades.
	Encargado de definir la mejor solución para un problema en particular en términos de: infraestructura (servidores, medios de almacenamiento, respaldo y recuperación, redes, etc.), tipo de DBMS a emplear. En general el arquitecto se enfoca a proporcionar la mejor solución	Generalmente son personas con una amplia trayectoria y extensa experiencia en proporcionar soluciones de todo tipo y tamaño, desde pequeños sistemas web hasta sistemas de misión crítica.
Arquitecto de bases de datos.	haciendo un amplio énfasis en los requerimientos funcionales, claro sin dejar de considerar los no funcionales.	
Consultor de bases de datos	Ofrece asesoramiento y consultoría a empresas en cuanto al uso de tecnologías de bases de datos con la finalidad de mejorar sus procesos de negocio y alcanzar metas específicas.	Todas las habilidades mencionadas anteriormente.



12. OBJETIVOS DE UN SGBD

1. No redundancia:

Los datos no deben estar duplicados

→

Gestión de accesos concurrentes

2. Consistencia (Integridad semántica): Evitar errores humanos

Los datos deben ser consistentes (sin fallos lógicos)

→

Mecanismos de mantenimiento de integridad

3. Fiabilidad: Control y recuperación de errores

Los datos deben estar protegidos contra fallos catastróficos

→

Mecanismos de mantenimiento de recuperación y relanzamiento de transacciones



Redundancia

- Ocurre cuando un mismo dato se almacena más de una vez de manera innecesaria en diferentes lugares.
- Puede provocar inconsistencia de datos

Inconsistencia de datos

- Ocurre cuando existen 2 o más versiones de un mismo dato cuyo valor es diferente (versiones en conflicto)



Ejemplo

Suponer que una empresa que ofrece préstamos a través de un sistema de caja de ahorro almacena los datos de sus clientes y sus préstamos de la siguiente manera:

Num. cliente	Nombre	Ap Paterno	Ap Materno	email	RFC	Folio préstamo	Fecha préstamo	Importe préstamo
1	Juan	Lara	Mora	j@m.com	LAJ800911L14	9090	10/01/19	6,500
1	Juan	Lira	Mora	jm@m.com	LAJ800911L14	9091	10/02/20	7,400
2	Rubén	López	Gil	gil@m.com	LOGI880203ZI	5589	11/03/20	8,500
3	Rubén	Mtz	Gil	ru@m.com	MARU891203Z	5589	11/03/20	8,500

De la tabla de datos que inconsistencias se pueden observar

A los datos que muestran inconsistencia se les conoce como datos que carecen de **Integridad**.

Integridad de datos

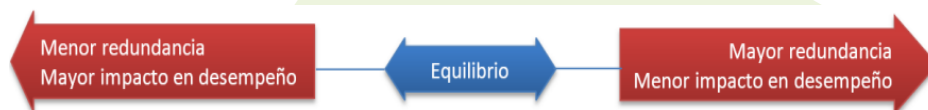


- Es una condición en la que todos los datos de una base se consideran consistentes con respecto al mundo real, a su contexto.
- En una BD con integridad los datos se consideran:
 - Adecuados (no existen inconsistencias).
 - Verídicos (los datos conducen a resultados consistentes).

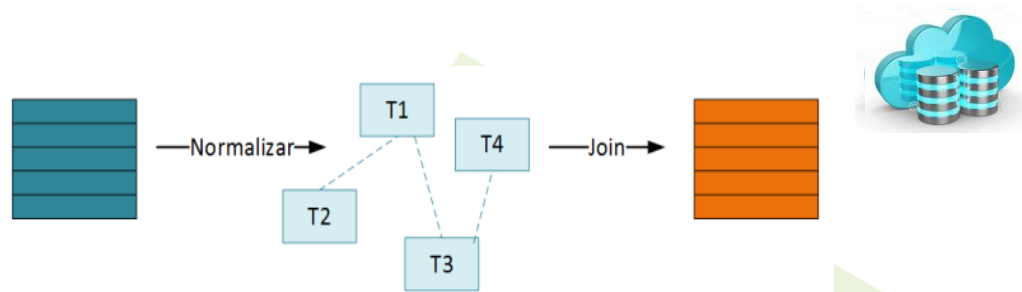
La redundancia no siempre es mala

- Es importante mencionar que en algunas situaciones se permite la existencia de cierto nivel de redundancia, siempre y cuando se cuente con un control adecuado para evitar la aparición de inconsistencias en los datos.
- ¿Por qué puede existir cierto nivel de redundancia?

Básicamente por desempeño, en una base de datos relacional, se pueden generar las siguientes situaciones:



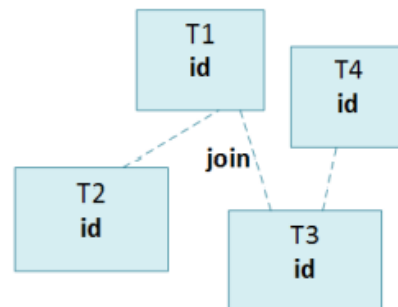
- Esta situación es especialmente particular para el modelo relacional.
- A nivel general, la estrategia para eliminar la redundancia de datos en el modelo relacional es a través de la división o separación de los datos que permitan eliminar duplicados.
- Lo anterior implica crear nuevas tablas que deberán estar relacionadas entre sí.



Redundancia necesaria

- Al realizar un proceso de división de los datos (normalización) para eliminar redundancias, los datos se distribuyen en nuevas tablas las cuales deben relacionarse entre sí.
- Estas relaciones permiten realizar la unión de los datos (join) cuando así se requiera.

- En el modelo relacional típicamente se incluye una columna cuyos valores permiten relacionar datos de una tabla con los datos de la otra (los clásicos ids).
- Podría decirse que esta columna es redundante en cada tabla ya que sus valores se repiten en 2 lugares diferentes.



Redundancia Innecesaria

- Corresponde a situaciones donde se cumple el concepto de redundancia visto anteriormente en las que su ocurrencia no se justifica.
- En estos casos, la redundancia puede eliminarse o conservarse dependiendo del punto de equilibrio. El nivel de normalización que permita encontrar este equilibrio depende de las características y requerimientos particulares de cada caso de estudio.

Practica 2:

Revisar la siguiente tabla de datos que representan la información de los proyectos y sus administradores de una empresa almacenados en archivos (hojas de cálculo). A cada proyecto se le asigna un solo administrador y un presupuesto.



registro	codigo_proyecto	administrador_proyecto	telefono_administrador	direccion_administrador	presupuesto_proyecto
1	21-5Z	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	16833460
2	21-5Z	William K. Moor	904-445-2719	216 Morton Rd., Stetson, FL 30155	16833460
3	25-2D	Jane D. Grant	615-898-9909	218 Clark Blvd., Nashville, TN 36362	12500000
4	25-5A	George F. Dort	615-227-1245	124 River Dr., Franklin, TN 29185	32512420
5	25-9T	Holly B. Parker	904-338-3916	3334 Lee Rd., Gainesville, FL 37123	21563234
6	27-4Q	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	10314545
7	29-2D	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	25559999
8	31-7P	William K. Moor	904-445-2719	206 Morton Rd., Stetson, FL 30155	56850000

- A. ¿Qué anomalías existen en los datos relacionados con el concepto de inconsistencia, redundancia e integridad?
- B. Revisar cada una de las columnas, indicar si la columna presenta redundancia de ser así indicar si es redundancia necesaria o innecesaria.
- C. Suponga que al administrador de los archivos le solicitan un reporte ordenado por la ciudad a la que pertenece cada administrador de proyecto, ¿Qué cambios aplicaría a la estructura de datos para poder satisfacer el reporte solicitado en el punto anterior?
- D. Una solución simple y básica para eliminar redundancia es la separación de los datos en varios archivos, ¿cómo podría ser la separación de archivos?



¿Preguntas????