

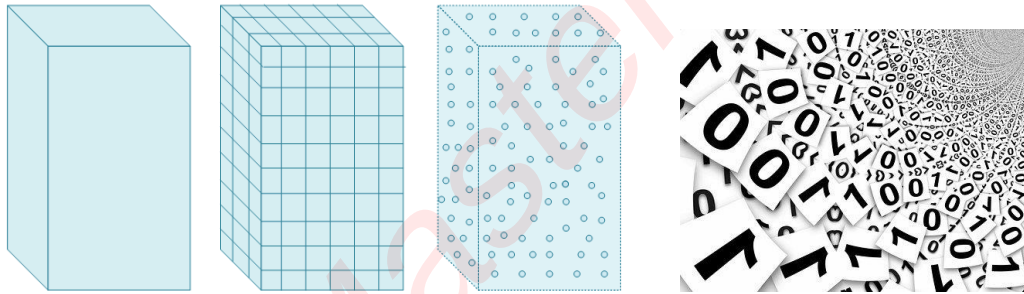
TEMA 1  
SISTEMAS DE BASES DE DATOS

1.1. 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.

1PB = 1024 TB = 1,048,576 GB

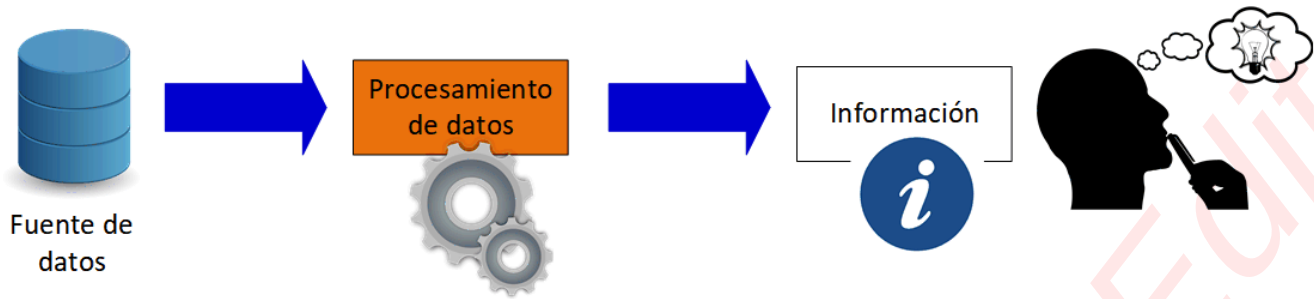


1.1.1. Dato.

- Es un símbolo (número, letra, signo ortográfico, etc.) empleado para representar a través de un lenguaje algún hecho, condición valor, situación, cantidad, o medida.
- Conjunto de símbolos en bruto (raw data) obtenidos a través de distintos medios como instrumentos, encuestas, mediciones, etc.

1.1.2. Información

- ¿Qué significado o interpretación tiene un dato?
  - Ninguno. Por sí solos los datos no tienen capacidad de comunicar un significado.
- ¿Qué sucede si a un conjunto de datos se les aplica un **procesamiento**?
  - Este procesamiento se realiza a través de equipos de cómputo de forma automatizada y a grandes velocidades.
  - El procesamiento puede ser tan simple o tan complejo como se requiera.
  - Como resultado de este procesamiento se obtiene **Información**.



Por lo anterior, la información representa el resultado obtenido del procesamiento de un conjunto de datos organizados y representados de forma correcta que revela o muestra un **significado**.

#### 1.1.2.1. Características de la información

- Significado (semántica)
- Importancia (relatividad al receptor)
- Vigencia (en la dimensión espacio-tiempo)
- Validez (relativa al emisor)
- Valor (activo intangible volátil)

#### 1.1.2.2. Utilidad:

- Descubrir el significado: ¿Qué nos quieren decir todos estos símbolos?
- Toma de decisiones. De vital importancia inclusive para la **supervivencia humana**

La obtención o generación de la información puede realizarse en 3 principales etapas:

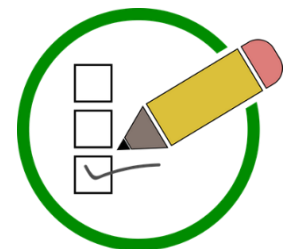
- Entrada o captura de datos.
- Representación y almacenamiento estructurado de datos en bruto (raw data).
- Representación e interpretación del resultado del procesamiento de datos.

#### Ejemplo:

Encuesta de evaluación de un profesor.

#### A. Entrada de datos (Input Data)

ENCUESTA			
<b>1. El interés del profesor para que los alumnos aprendan es:</b>			
<input type="radio"/> Muy grande	<input type="radio"/> Grande	<input checked="" type="radio"/> Escaso	<input type="radio"/> Nulo
<b>2. La confianza que inspira el profesor es:</b>			
<input type="radio"/> Muy grande	<input type="radio"/> Grande	<input checked="" type="radio"/> Escaso	<input type="radio"/> Nulo
<b>3. Las clases se desarrollan de manera:</b>			
<input type="radio"/> Muy interesante	<input type="radio"/> Interesante	<input checked="" type="radio"/> Aburrida	<input type="radio"/> Muy aburrida
<b>4. El profesor expone la clase de forma:</b>			
<input type="radio"/> Siempre	<input type="radio"/> Frecuentemente	<input type="radio"/> Algunas veces	<input checked="" type="radio"/> Nunca
<b>5. Los conocimientos del profesor sobre la asignatura son:</b>			
<input type="radio"/> Excelentes	<input type="radio"/> Buenos	<input type="radio"/> Regulares	<input checked="" type="radio"/> Nulos



#### B. Almacenamiento de datos en bruto (Raw Data) sin significado.

CLAVE_PROF	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
1009	A	A	B	A	A	B	A	A	A	B	A	A	A
1010	B	A	B	A	C	A	B	A	A	A	A	A	A
1090	B	C	B	C	B	A	C	C	B	A	C	B	C
8913	D	D	D	C	C	D	D	C	C	D	D	D	D

C. Obtención de información en diferentes formatos. Los datos adquieren significado después de su procesamiento.

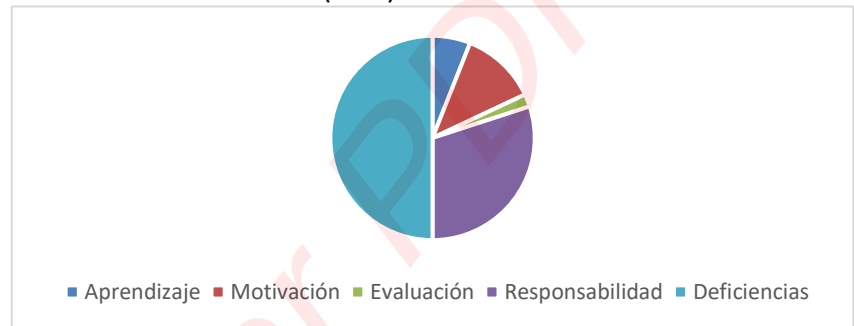
**Prof.** Juan López López.

#### Puntaje

Aprendizaje	6 de 30
Motivación	12 de 32
Evaluación	2 de 18
Responsabilidad	4 de 20
Promedio general	37%

**Encuestados:** 14 de 40 (28%)

**Promedio Global: 48.3**



Resultados por pregunta.

Num.	Pregunta	Calif	Max.	Min	Prom
1	Interés para qué los alumnos aprendan.	20	96	18	81.3
2	Confianza que inspira a los alumnos para intervenir en clase.	50	96	48	80.67
3	Las clases se desarrollan de manera interesante	60	88	50	76.5
4	Expone la clase con claridad.	56	96	40	78.3
5	Conocimientos sobre la asignatura.	30	100	30	87.7

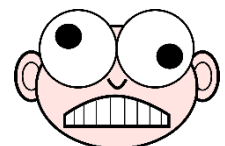
El resultado del procesamiento de la tabla de datos, genera información como la anterior. Con estos resultados, por ejemplo, es posible realizar **toma decisiones** acerca de este profesor. Como es de imaginarse, el profesor tendrá algunos problemas para permanecer en su puesto. Se llega a esta conclusión por sus bajas calificaciones, las cuales fueron calculadas a través del **procesamiento de los datos** de su encuesta.

#### 1.2. ADMINISTRACIÓN DE LOS DATOS:

Para poder 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?

CLAVE_PROF	P1	P2	P3	P4	P5	P6	P7
1009-0	OK	A		-2	A	1	V
JUAN LOPEZ	true	false	D		FALSE	0	F
8913	4	10	1		0	1	F
1009-0	OK	B			A	1	V



- 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.
- Respaldos
- Seguridad
- Interpretación de lenguajes de acceso a datos, interfaces de programación
- Interfaces de comunicación: Local, remota, etc.

Para poder 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*.

### 1.3. BASES DE DATOS Y SISTEMA ADMINISTRADOR DE BASES DE DATOS (DBMS).

#### 1.3.1. Bases 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:



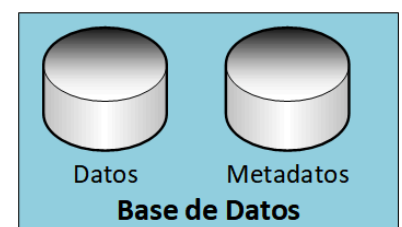
- Notar que el concepto de Bases de Datos, es diferente al concepto de DBMS.
- La colección de datos almacenados es administrada por un DBMS, y se divide en 2 grupos:

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



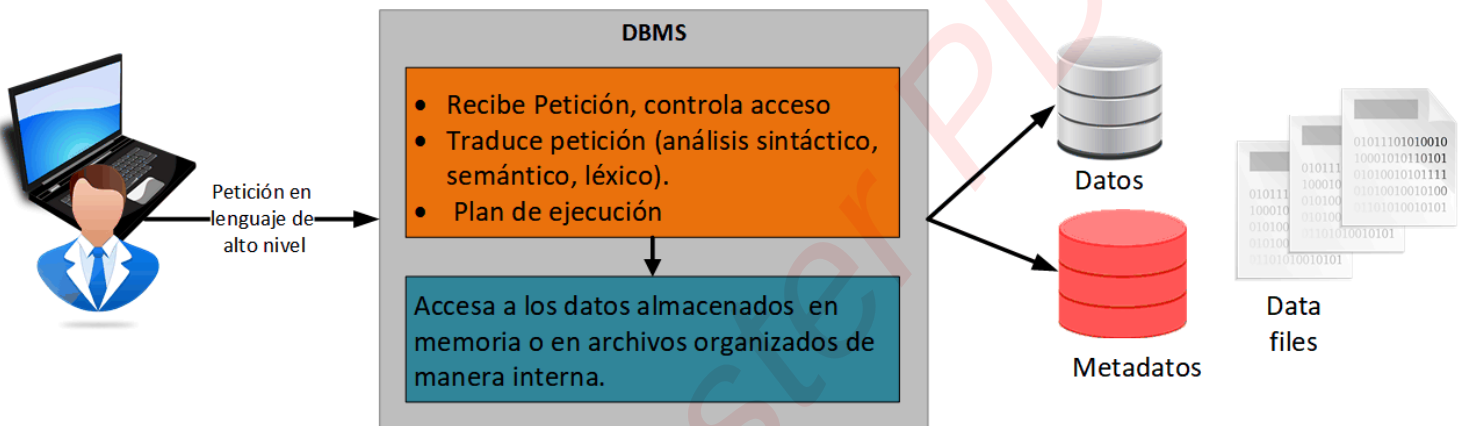
### 1.3.2. Database Management System (DBMS).

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

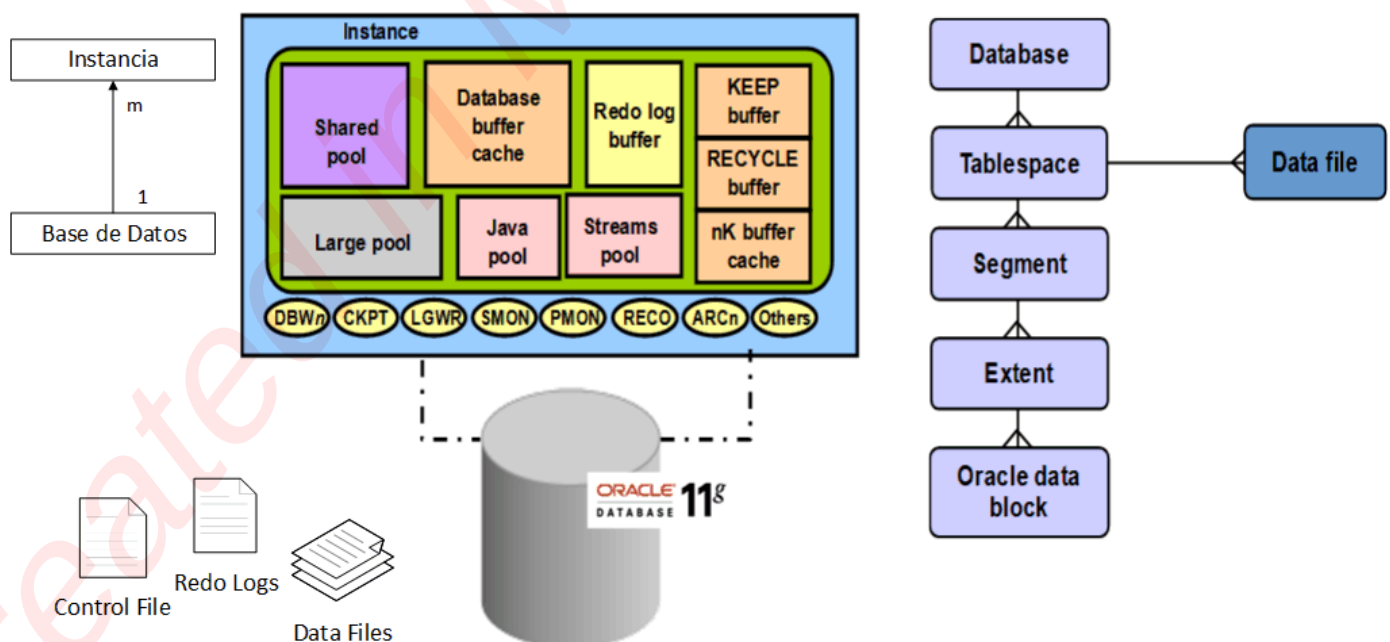
#### Principales funciones:

- 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**.

### 1.3.3. Arquitectura básica de un DBMS



#### 1.3.3.1. Ejemplo, arquitectura Oracle 11g



#### 1.4. TIPOS DE BASES DE DATOS.

Existen diversas clasificaciones de las bases de datos. En realidad, no existe un estándar o regla general para clasificarlas. 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**



En el siguiente capítulo se revisarán algunos de estos tipos de bases de datos y sus respectivos modelos.

#### 1.5. INTEGRIDAD, REDUNDANCIA Y CONSISTENCIA

##### 1.5.1. Redundancia.

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

##### 1.5.2. 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	LOGI880203Z1	5589	11/03/20	8,500
3	Rubén	Martínez	Gil	ru@m.com	MARU891203Z	5589	11/03/20	8,500

De la tabla de datos se pueden observar las siguientes inconsistencias:

- Los datos del email del cliente 1 son inconsistentes. ¿Cuál es el correcto?
- El apellido paterno del cliente 1 es inconsistente.
- Observar que los clientes 2 y 3 son 2 personas diferentes y están asociados al mismo préstamo. En teoría cada préstamo debe ser solicitado por un solo cliente.



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

### 1.5.3. 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.
- Una BD con integridad los datos se consideran:
  - Adecuados (no existen inconsistencias).
  - Verídicos (los datos conducen a resultados consistentes).

### 1.5.4. 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?

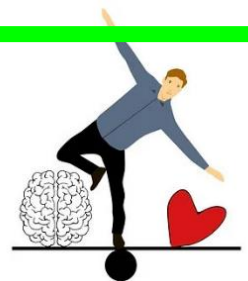
Básicamente por **desempeño**. Como se verá en capítulos posteriores, en una base de datos relacional, se pueden generar las siguientes situaciones:

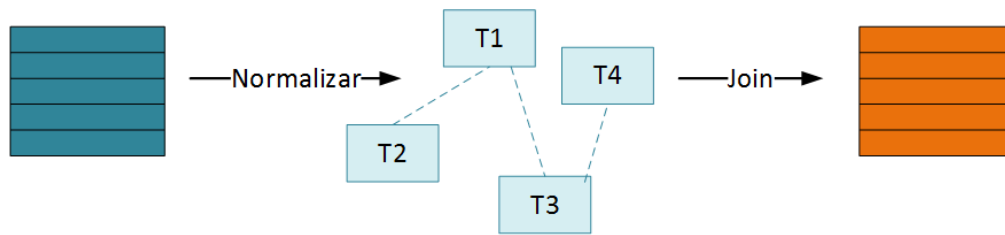
Menor redundancia  
Mayor impacto en desempeño

Equilibrio

Mayor redundancia  
Menor impacto en desempeño

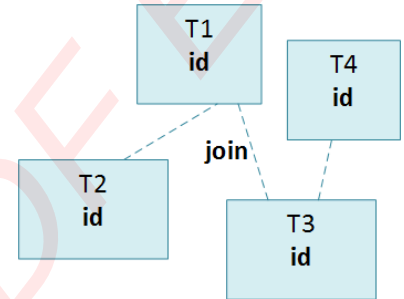
- Esta situación es especialmente particular para el modelo relacional.
- A nivel general, la estrategia para eliminar la redundancia de datos en el modelo relacional es a través de la **división** o separación de los datos que permitan eliminar duplicados.
- Lo anterior implica crear nuevas tablas que deberán estar relacionadas entre sí.
- Esta división implica que la recuperación o consulta de datos será más costosa ya que se deberá ahora hacer la operación contraria, es decir, una operación **join** que permitirá unir nuevamente a los datos para ser presentados como si estuvieran almacenados en una sola tabla de datos.





#### 1.5.4.1. 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**. Sin embargo su uso es totalmente necesario ya que permite relacionar a los datos que han sido normalizados en diferentes tablas.



#### 1.5.4.2. 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.

#### Ejemplo:

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?

Respuesta:



- I. El proyecto 21-5Z tiene 2 administradores.
- II. El nombre George F. Dorts aparece diferente en algunos registros.
- III. El teléfono de Holly B. Parker aparece de formas diferentes en algunos registros.
- IV. El número de casa de William K. Moor es inconsistente en algunos registros.

B. Revisar cada una de las columnas, indicar si la columna presenta redundancia de ser así indicar si es redundancia necesaria o innecesaria.

Respuesta:

Nombre de la columna	Resultado
registro	No redundante
codigo_proyecto	Redundancia innecesaria
administrador_proyecto	Redundancia innecesaria
teléfono_administrador	Redundancia innecesaria
dirección_administrador	Redundancia innecesaria
presupuesto_proyecto	Redundancia innecesaria

- No existe redundancia necesaria ya que los datos aún no se han dividido.

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 solicitando en el punto anterior?

Respuesta:

Para satisfacer este requerimiento, se necesita dividir el campo DIRECCION\_ADMINISTRADOR en su forma más simple: calle, número, ciudad, etc.

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?

Respuesta:

Los nuevos archivos contendrían la siguiente información (uno por tabla).

#### Administrador

clave_admin	nombre	telefono	calle	numero	colonia	ciudad	codigo_postal
1	Holly B. Parker	904-338-3416	Lee Rd.	3334	Gainesville	, FL	37123
2	William K. Moor	904-445-2719	Morton Rd.	216	Stetson	FL	30155
3	Jane D. Grant	615-898-9909	Clark Blvd.	218	Nashville	TN	36362
4	George F. Dort	615-227-1245	River Dr.	124	Franklin	TN	29185

#### Proyecto

Código_proyecto	presupuesto	clave_admin
21-5Z	16833460	1
25-2D	12500000	3
25-5A	32512420	4
25-9T	21563234	1
27-4Q	10314545	4
29-2D	25559999	1
31-7P	56850000	2

- Observar que se crea una nueva columna “artificial” clave\_admin para poder asociar a ambas tablas.

- Esta columna es un ejemplo de una redundancia necesaria.
- Notar que la columna le pertenece a la tabla `admin` y se propaga o se copia hacia la tabla `proyecto`.

¿Qué pasaría si se en lugar de establecer en este sentido la relación, se hubiera optado por hacer uso de la columna `codigo_proyecto` y propagarla hacia la tabla `admin`?

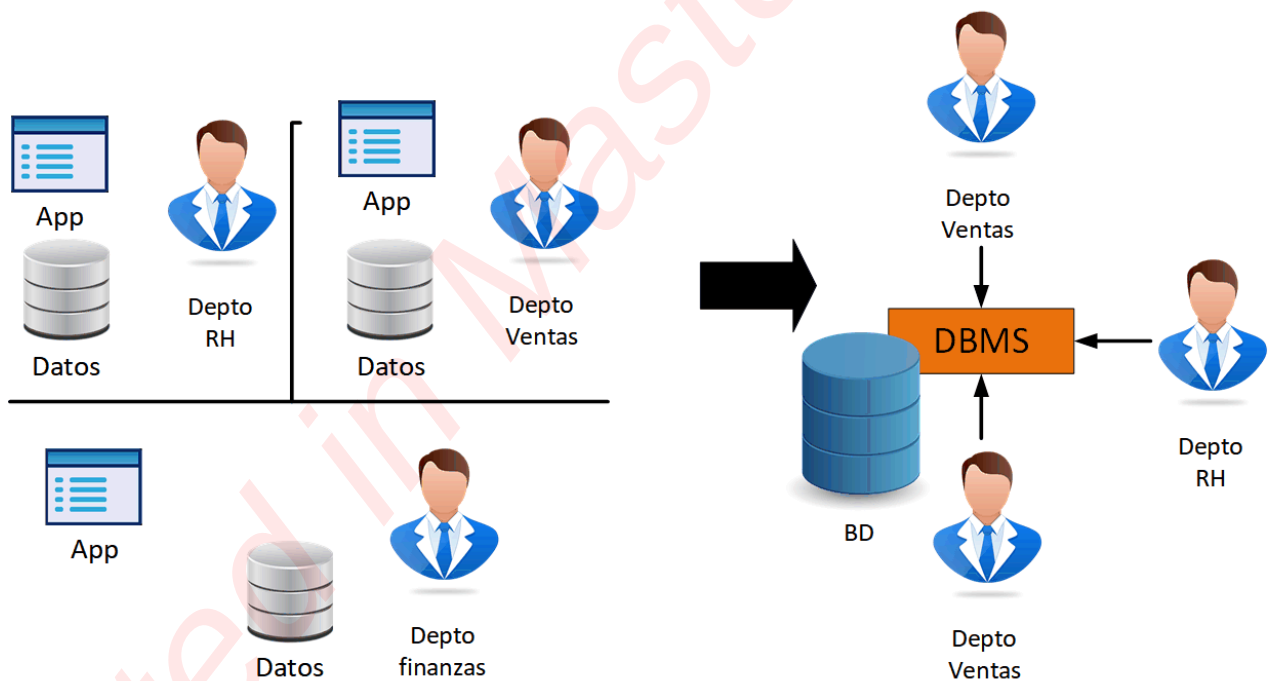
- Si se invierte la relación no se cumpliría la regla que establece que a cada proyecto se le debe asignar un solo administrador debido a que la presencia de la columna `codigo_proyecto` en la tabla `admin` podría tener valores duplicados y así un mismo proyecto podría estar asociado a distintos administradores.

#### Administrador

clave_admin	nombre	telefono	calle	numero	colonia	ciudad	codigo_postal	codigo_proyecto (Incorrecto)
1	Holly B. Parker	904-338-3416	Lee Rd.	3334	Gainesville	, FL	37123	21-5Z
2	William K. Moor	904-445-2719	Morton Rd.	216	Stetson	FL	30155	25-2D
3	Jane D. Grant	615-898-9909	Clark Blvd.	218	Nashville	TN	36362	25-5A
4	George F. Dort	615-227-1245	River Dr.	124	Franklin	TN	29185	25-5A

#### 1.6. SISTEMAS DE BASES DE DATOS.

El siguiente diagrama muestra la 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.




En un sistema de bases de datos, la administración de los datos se realiza a través del empleo de un DBMS. Los integrantes de la organización hacen uso de él en lugar de tener bases aisladas o sistemas de archivos particulares a cada departamento.


##### 1.6.1. Elementos de un sistema de bases de datos.

Un sistema de bases de datos está integrado por los siguientes elementos.

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

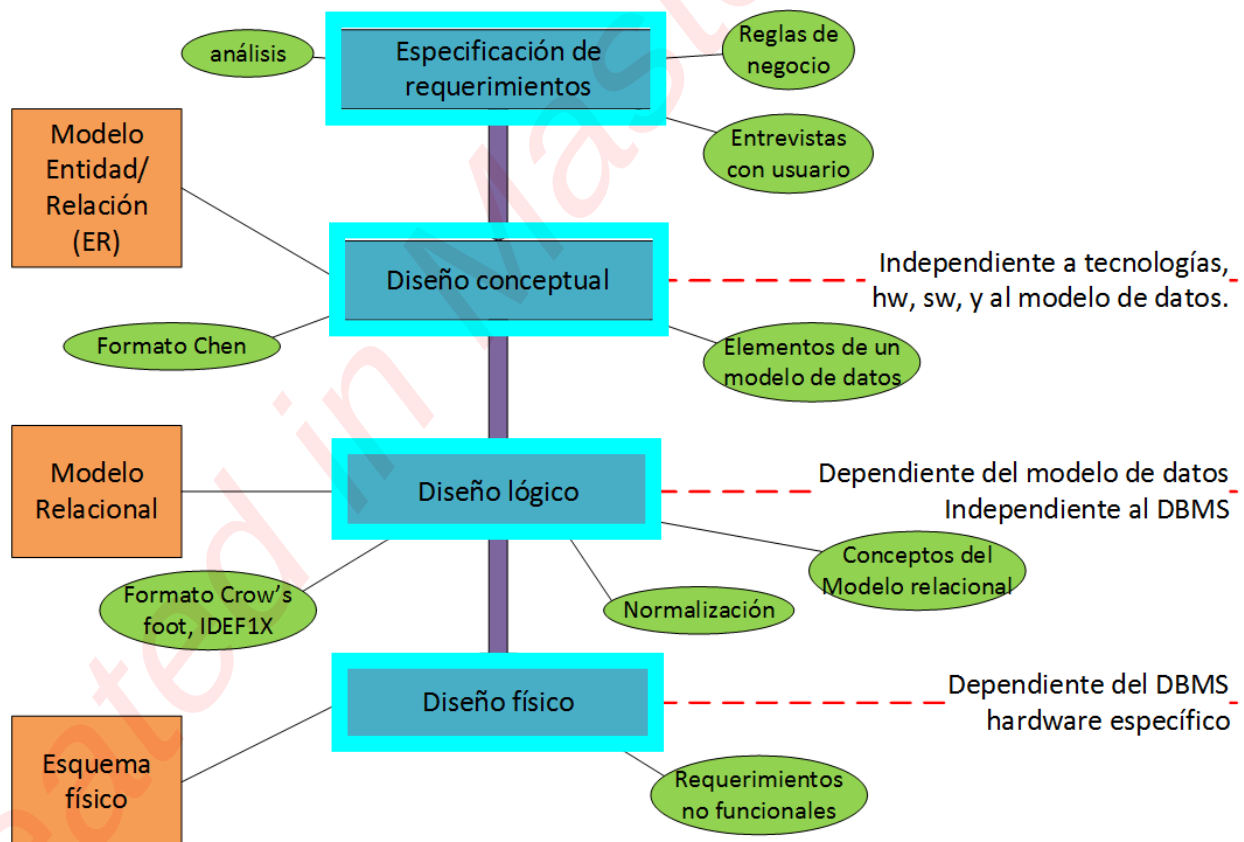
### 1.6.2. Roles asociados con bases de datos.

Nombre del rol	Descripción	Principales habilidades.
<b>Desarrollador</b> / Programador de bases de datos.	Crea y mantiene aplicaciones desarrolladas con herramientas, extensiones y lenguajes de programación de bases de datos.	<b>Programación SQL</b> , y extensiones PL/SQL, etc. 
<b>Analista</b> de bases de datos. 	Generación de los artefactos necesarios para extraer y comprender los <b>requerimientos tanto funcionales como no funcionales</b> de un usuario o cliente final.	Facilidad de <b>interacción con el usuario final</b> , 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.
<b>Diseñador</b> de la base de datos. 	A partir de un caso de estudio o problema real, generar diferentes <b>modelos de datos que describen la estructura</b> y el comportamiento de los datos tratando de conservar un equilibrio entre desempeño, niveles de normalización, etc.	<b>Diseño de modelos de datos</b> , conceptos sólidos de programación orientada a objetos, Ingeniería de software, SQL.
Administrador de la base de datos ( <b>DBA</b> ) 	Encargado de <b>mantener la buena salud del DBMS</b> para que este siempre opere de la forma más óptima posible.  <b>Tarea:</b> Leer Artículo Responsabilidades del DBA en tiempos 2.0. Obtener el artículo de: • <a href="http://dbagroup.cl/blog/?p=200">http://dbagroup.cl/blog/?p=200</a> 	Conocimientos sólidos en <b>arquitecturas y algoritmos internos</b> de la base de datos, estructuras <b>físicas y lógicas</b> tanto de almacenamiento como de <b>memoria y procesos</b> , además de contar con fundamentos sólidos en áreas como: <b>Sistemas operativos, redes, diseño y programación de sistemas de software, seguridad</b> , etc.
Especialista en <b>análisis de datos</b> .	Sus conocimientos le permiten aplicar tecnologías, técnicas, cálculos <b>matemáticos y estadísticos</b> , así como el uso de una diversidad de herramientas enfocadas al <b>análisis de datos</b> de grandes volúmenes. <b>Data warehouse</b> , <b>BigData</b>	Conocimientos sólidos en las áreas de <b>estadística, inteligencia artificial</b> , así como el uso de diversas herramientas y tecnologías enfocadas al análisis de datos que permitan <b>descubrir conocimiento</b> y con ello la toma de

		decisiones importantes para una empresa u organización.
<b>Arquitecto</b> de bases de datos. 	Encargado de definir la mejor solución para un problema en particular en términos de: <b>infraestructura</b> (servidores, medios de almacenamiento, respaldo y recuperación, redes, etc.), tipo de <b>DBMS</b> a emplear. En general el arquitecto se enfoca a proporcionar <b>la mejor solución</b> haciendo un amplio énfasis en los <b>requerimientos funcionales</b> , claro sin dejar de considerar los funcionales.	Generalmente son personas con una amplia <b>trayectoria y extensa experiencia</b> en proporcionar soluciones de todo tipo y tamaño, des de pequeños sistemas web hasta sistemas de misión crítica.
<b>Consultor</b> de bases de datos	Ofrece <b>asesoramiento y consultoría</b> a empresas en cuanto al uso de tecnologías de bases de datos con la finalidad de <b>mejorar sus procesos de negocio</b> y alcanzar metas específicas.	<b>Todas las habilidades mencionadas anteriormente.</b>

### 1.7. METODOLOGÍAS DE DISEÑO DE BASES DE DATOS.

Para realizar el correcto diseño de una base de datos se requiere del uso de una metodología que permite identificar las diferentes etapas, así como las diferentes herramientas a utilizar. Una de las principales metodologías se ilustra en el siguiente diagrama.



#### 1.7.1. Especificación de requerimientos

- Su objetivo es describir con precisión el contenido de la información de la base de datos
- Se identifican a los grupos de usuarios de la empresa u organización que van a interactuar con la BD.
- Para cada grupo se identifican sus requerimientos y especificaciones.
- Cada una de estas especificaciones representa una vista particular de la BD.
- Se identificarán los objetos o eventos del mundo real que almacenará la BD así como sus atributos y las relaciones entre ellos.
- Se identifican las condiciones de integridad, restricciones empleando como principal insumo las reglas de negocio.
- Se identifican autorizaciones de acceso a la información por parte de los diferentes grupos de usuarios.
- Importante: Realizar una estimación del volumen de datos que habrá de almacenarse en la base de datos a lo largo del tiempo.
- Finalmente, es altamente recomendable identificar las principales operaciones a realizar con los datos, es decir, la descripción de las transacciones: frecuencia, datos que utilizan y producen, así como el flujo de información entre ellas.

#### 1.7.2. *Diseño conceptual*

- En esta etapa se realiza la integración de las diferentes vistas obtenidas en la etapa anterior. Se genera una descripción global de la BD eliminando inconsistencias y redundancia entre cada vista.
- Se realiza la construcción de un modelo de datos que describe a cada uno de los objetos y relaciones identificados en la etapa anterior.
- Por lo general se emplea el Modelo Entidad –Relación propuesto por Peter Chen.
- Este modelo es totalmente **independiente** al modelo de datos a emplear (red, Jerárquico, relacional, al DBMS que se va a emplear para implementar la base de datos y al hardware que se empleará.
- A dicho modelo se le conoce como modelo o esquema conceptual, modelo Entidad-Relación

#### 1.7.3. *Diseño lógico.*

- En esta etapa el modelo conceptual se transforma en el modelo empleado por el DBMS. Para el curso se trata del modelo relacional. A este nuevo modelo se le conoce como Modelo o diagrama relacional.
- Existen diversas notaciones para representar este modelo. Las notaciones Crow's foot e IDEF1X son las que se emplearán en el curso.
- El modelo generado es **dependiente** del modelo de datos a emplear (en este caso al relacional), e **independiente** al DBMS (Oracle, MySQL, PostgreSQL, SQL Server, etc.).
- Para el caso del modelo relacional, en esta etapa se construyen todos los elementos necesarios para representar a los elementos del modelo conceptual, como son tablas, restricciones, etc.
- Se crean los usuarios en el DBMS así como sus reglas de acceso a los datos (privilegios).
- En algunas ocasiones el modelo relacional se somete a un proceso en el que se identifican anomalías en especial de redundancia que pueden ser eliminadas a través de un proceso llamado Normalización.

#### 1.7.4. *Diseño físico.*

- A partir del diseño lógico y de una estimación de uso, volumen, concurrencia en los datos, se crea una **configuración física** que será adaptada al entorno donde la BD será instalada. Esta configuración permitirá el almacenamiento y la explotación de datos con el mejor rendimiento posible.

- La configuración física incluye principalmente a los componentes de hardware como son la creación de clusters, particionamiento, distribución de discos, redes de datos, etc.
- Este diseño es totalmente **dependiente** del DBMS a emplear.