

Introducción a Bases de Datos

Importancia de las Bases de Datos

El procesamiento de la información es fundamental para planificar cualquier actividad del hombre de hoy, desde conocer el tiempo, informarse sobre cotizaciones de la bolsa o el conocimiento actualizado de datos de la organización.

Una consecuencia de todo esto es la gran cantidad de datos que es necesario manejar en cualquier actividad de la vida. Al aumentar el volumen de datos a manejar, también es necesario aumentar la rapidez con que se procesan dichos datos.

Antes de la aparición del concepto de Base de Datos, la información se almacenaba en archivos sin ninguna estructura, denominados archivos planos.

Al final de los años 60, para paliar los defectos de los sistemas de información existentes surge un nuevo concepto que fue denominado Base de Datos.

¿Por qué utilizar Bases de Datos?

El concepto de Bases de Datos, tiene como principal característica la integración de distintos archivos individuales en una única Base de Datos, donde la información es compartida por todos los usuarios.

El concepto de Base de Datos, surge como propuesta de solución a un conjunto de problemas técnicos y administrativos presentes en el manejo de archivos. A medida que los sistemas de información se volvían más complejos y se extendían a nuevas áreas de operación de una organización las dificultades en mantener el funcionamiento y control se acentuaban.

Problemas de los sistemas tradicionales

Redundancia e inconsistencia de los datos:

Este problema origina diversos archivos con diferentes formatos y la misma información puede estar duplicada en distintos archivos. Esta redundancia conduce a un almacenamiento y coste de acceso alto, además puede conducir a inconsistencia de datos (diversas copias de los mismos datos pueden no coincidir).

Dificultades en el acceso a los datos.

El entorno de procesamiento de archivos convencional no permite que los datos sean obtenidos de forma práctica y eficiente.

Aislamiento de los datos.

Debido a que los datos están dispersos en varios archivos, y los archivos están en diferentes formatos, es difícil escribir nuevos programas de aplicación para recuperar los datos.

Problemas de integridad

Los valores de los datos almacenados deben satisfacer ciertos tipos de restricciones, cuando se añaden nuevas restricciones, es difícil cambiar los programas para hacer que se cumplan.

El problema se complica aún más cuando las restricciones implican distintos elementos de datos de diferentes archivos.

Problemas de atomicidad

En muchas aplicaciones es crucial asegurar que una vez que un fallo ha ocurrido y se ha detectado, los datos se restauran al estado de consistencia que existía antes del fallo. Es muy difícil asegurar esta propiedad en un sistema de procesamiento de archivos convencional.

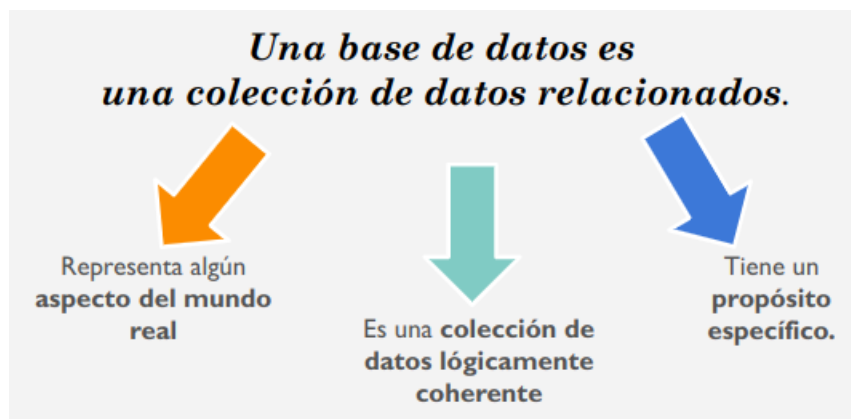
Anomalías de acceso concurrente

En los sistemas tradicionales la interacción de actualizaciones concurrentes puede dar lugar a datos inconsistentes ya que la supervisión necesaria es difícil de lograr. Este problema no permite a múltiples usuarios actualizar los datos simultáneamente.

Problemas de seguridad

No todos los usuarios o aplicaciones deberían poder acceder a todos los datos.

Definiciones de Bases de Datos



“Una Base de Datos es un conjunto de datos almacenados de forma integrada y compartida.”

Sistemas de Gestión de Bases de Datos DBMS (Database Management System)

Las dificultades de los sistemas tradicionales dieron origen al desarrollo de SGBD (Sistemas de Gestión de Bases de Datos).

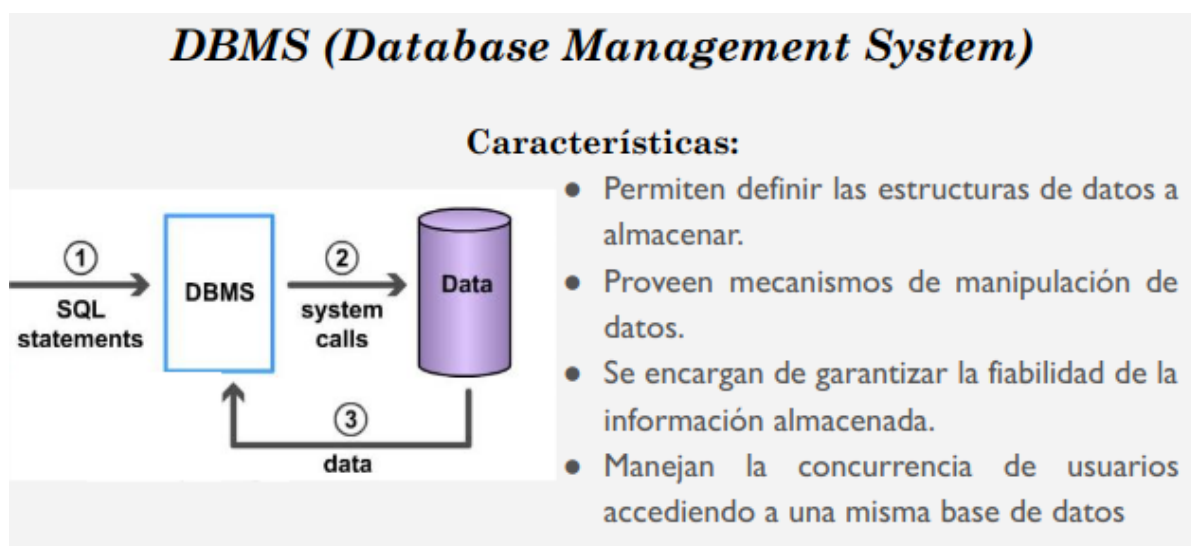
Este enfoque de Bases de Datos se caracteriza por estar orientado a los datos, siendo estos un conjunto estructurado e independiente de las aplicaciones; donde su objetivo es satisfacer las necesidades de información de las aplicaciones.

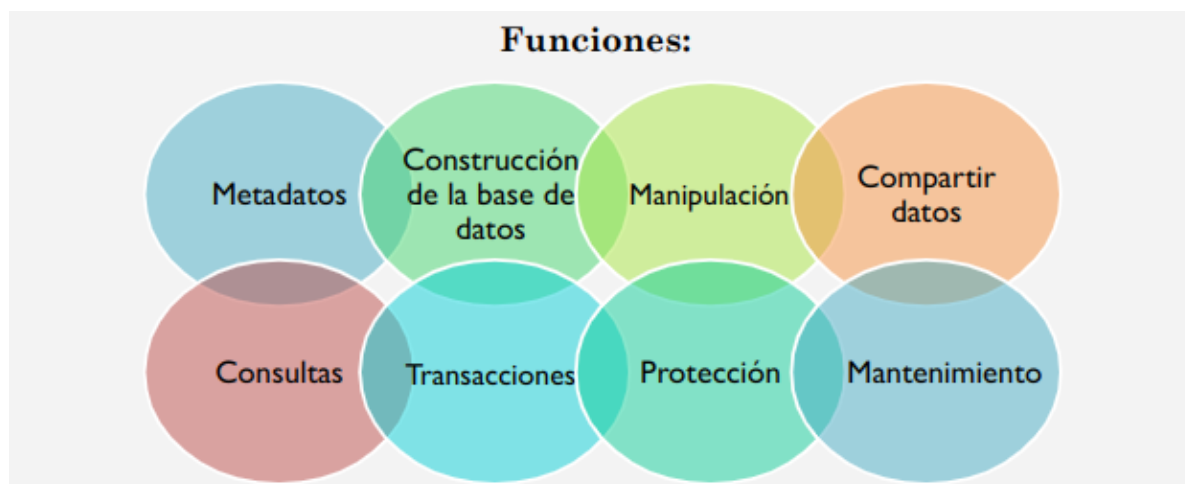
Un sistema de Bases de Datos es un sistema mecanizado por computadoras para el manejo de datos por medio de paquetes de software llamados Sistemas de Gestión de Bases de Datos o DBMS (Database Management System).

Un sistema de gestión de bases de datos es una colección de archivos interrelacionados y un conjunto de programas que permiten a los usuarios acceder, introducir, almacenar, modificar, eliminar y ordenar los datos.

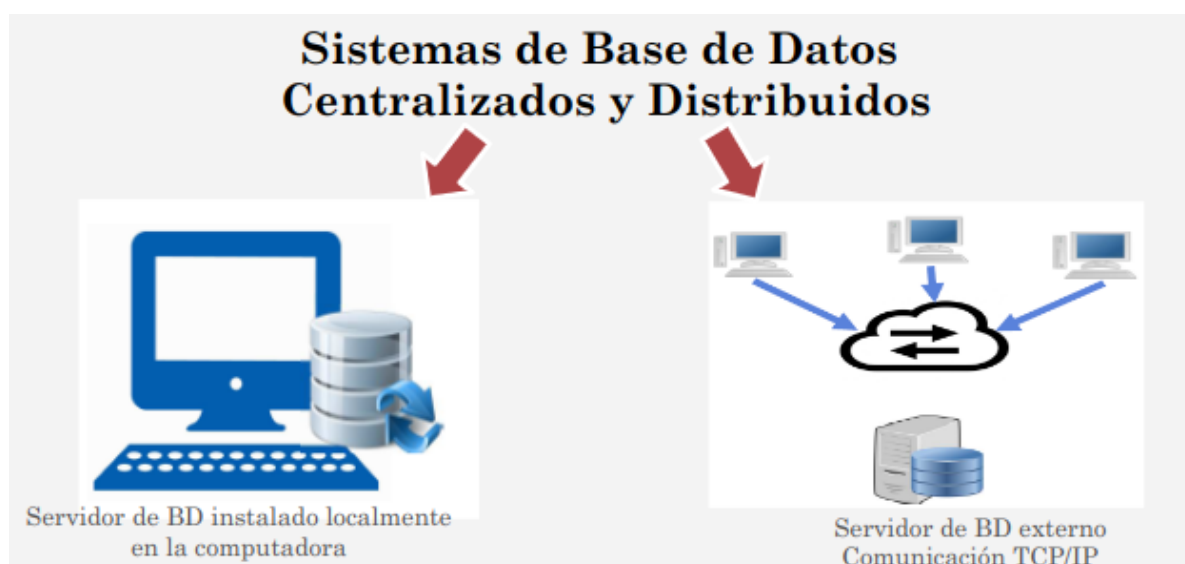
Las principales funciones de un DBMS son:

- Crear y organizar la Base de Datos.
- Permitir introducir, almacenar y recuperar datos de la BD.
- Establecer y mantener las rutas de acceso a cualquier dato de la Base de Datos de la forma más rápida posible.
- Manejar los datos según las necesidades de los usuarios.
- Mantener la integridad y seguridad de los datos.





Tipos de Sistemas de Gestión de Bases de Datos



Ventajas de los Sistemas de Gestión de Bases de Datos

Los sistemas de bases de datos presentan numerosas ventajas.

Permiten:

- Controlar la redundancia de datos.
- Controlar la consistencia de los datos.
- Compartir los datos.
- Mantener estándares.
- Mejorar la integridad de los datos.
- Mejorar el mantenimiento.
- Aumentar la concurrencia.
- Mejorar la seguridad.

La integración de los datos y la utilización del SGBD plantean los siguientes inconvenientes:

- Alta complejidad.
- Gran tamaño.
- Costo del SGBD.
- Costo del equipamiento.
- Vulnerable a fallos.

Personas en el entorno de una de Bases de Datos

Podemos identificar cuatro grupos de personas que intervienen en el entorno de una base de datos:

Administrador:

- Se encarga de la implementación física de la base de datos.
- Toma decisiones relativas al almacenamiento físico según las posibilidades que le ofrezca el SGBD.
- Es responsable de establecer la política de seguridad.
- Debe monitorear que el sistema se encuentre siempre operativo y procurar que los usuarios y las aplicaciones obtengan buenas prestaciones.

Diseñadores:

- Realizan el diseño de la base de datos, deben identificar los datos, las relaciones entre ellos y las restricciones sobre los datos y sobre sus relaciones.
- Deben tener un profundo conocimiento de los datos de la organización y también deben conocer sus reglas de negocio.

Programadores:

- Realizan la implementación de los programas de aplicación que utilizan los usuarios finales.
- Los programas de aplicación son los que permiten consultar, insertar, actualizar y eliminar los datos almacenados en la BD.

Usuarios finales:

- Representan a los clientes de la base de datos.
- La base de datos se diseña, implementa y mantiene, para satisfacer sus requisitos en la gestión de la información.

Esquema de una de Bases de Datos

El diseño completo de la Base de Datos se denomina esquema, y estos raramente son modificados.

Definimos como esquema de la Base de Datos a la estructura de la Base de Datos, es decir su forma, su esqueleto.

Los sistemas de Bases de Datos tienen varios esquemas divididos de acuerdo a los niveles de abstracción.

- Nivel más bajo: esquema físico
- Nivel intermedio: esquema lógico
- Nivel más alto: subesquema.

En general, los sistemas de Bases de Datos soportan un esquema físico, un esquema lógico y varios subesquemas.



Para reducir al usuario la complejidad, las Bases de datos utilizan distintos niveles de abstracción.

Nivel Físico: es el nivel más bajo de abstracción, describe cómo se almacenan realmente los datos.

Nivel Lógico: describe qué datos se almacenan en la Base de Datos y qué relaciones existen entre esos datos.

Nivel de vistas: este es el nivel de abstracción más alto donde solo se describe parte de la Base de Datos completa.

Independencia de los Datos

Definimos como independencia de los datos a la capacidad de modificar una definición de un esquema sin afectar al esquema de nivel superior.

Un aspecto muy importante a considerar es:

- Cómo afecta la modificación de un esquema de un nivel, a la definición de los esquemas o subesquemas del nivel superior.

Independencia física: es la capacidad para modificar el esquema físico sin provocar que los programas de aplicación tengan que reescribirse.

Independencia lógica: es la capacidad de modificar el esquema lógico sin causar que los programas de aplicación tengan que reescribirse.

¿Cuál de las dos independencias es la más difícil de lograr?

La independencia lógica de los datos es la más difícil de lograr. Esto ocurre porque los programas de aplicación son fuertemente dependientes de la estructura lógica de los datos a los que acceden.

Modelos de Datos

¿Qué significa DATO?

Dato es el conjunto de propiedades que caracterizan un fenómeno. Denominamos información al conjunto de valores y relaciones entre estas propiedades (datos).

¿Qué significa MODELAR?

Modelar consiste en definir un mundo abstracto y teórico, tal que las conclusiones que se puedan sacar de él coincidan con las manifestaciones aparentes del mundo real [Flory, 1982].

Los modelos de datos son el instrumento principal para ofrecer la abstracción a través de su jerarquía de niveles. Un modelo de datos es un conjunto de conceptos que sirven para describir la estructura de una base de datos. Es decir, describir los datos, las relaciones entre los datos y las restricciones que deben cumplirse sobre los datos.

Razones para separar el dato de su significado:

- Las computadoras no manejan (bien) el lenguaje natural, que es la mejor forma de dar interpretación y significado a un dato.
- El almacenamiento del significado de los datos ocupa espacio.

En el lenguaje natural cada dato y su significado se registran juntos, ya que el lenguaje es lo suficientemente poderoso para hacerlo.

Pero para las computadoras es necesario separar el dato y la semántica ya que no manejan bien el lenguaje natural.

- Los datos no son estáticos ya que el mundo está en constante cambio.
- Los modelos de datos nos permiten capturar parcialmente el mundo, ya que es improbable generar un modelo que lo capture totalmente.

Se utilizan modelos de datos para describir un esquema, estructura.

Define las reglas por las cuales los datos son estructurados.

Define las operaciones que se permiten efectuar con los datos. Generalmente están relacionadas con la estructura de los datos y tienen validez en el contexto en que fueron definidas.

Todo modelo de datos debe poder capturar las propiedades estáticas y dinámicas de una realidad.

- Propiedades estáticas son aquellas que no varían en el tiempo. Quedan definidas en el modelo de datos por las estructuras. Un ejemplo de ellas es que los precios se miden en pesos, por ejemplo.
- Propiedades dinámicas son aquellas que varían, evolucionan en el tiempo.

Un Modelo de Datos consta de:

- Un conjunto de reglas que lo generan
- Un conjunto de operaciones

Existen distintos tipos de modelos de acuerdo a los niveles de detalle que se requieran y también al nivel de abstracción en el cual se esté trabajando.

El modelo de datos es una parte esencial de la estructura de una Base de Datos.

Es una colección de herramientas conceptuales para describir los datos, las relaciones entre ellos, la semántica de los datos y las restricciones de consistencia.

Los distintos modelos propuestos pueden ser agrupados en tres grupos: modelos lógicos basados en objetos, modelos lógicos basados en registros y modelos físicos.

Modelos lógicos basados en objetos

- Describen los datos en los niveles de abstracción de vistas y lógico.
- Brindan capacidades estructurales muy flexibles.
- Permiten especificar explícitamente las restricciones.

Los modelos de este tipo más comúnmente empleados son:

- Modelo entidad relación
- Modelo orientado a objetos
- Modelo de datos semánticos
- Modelo de datos funcional.

Cronología de los Modelos de Datos

- Basados en sistemas de archivos convencionales

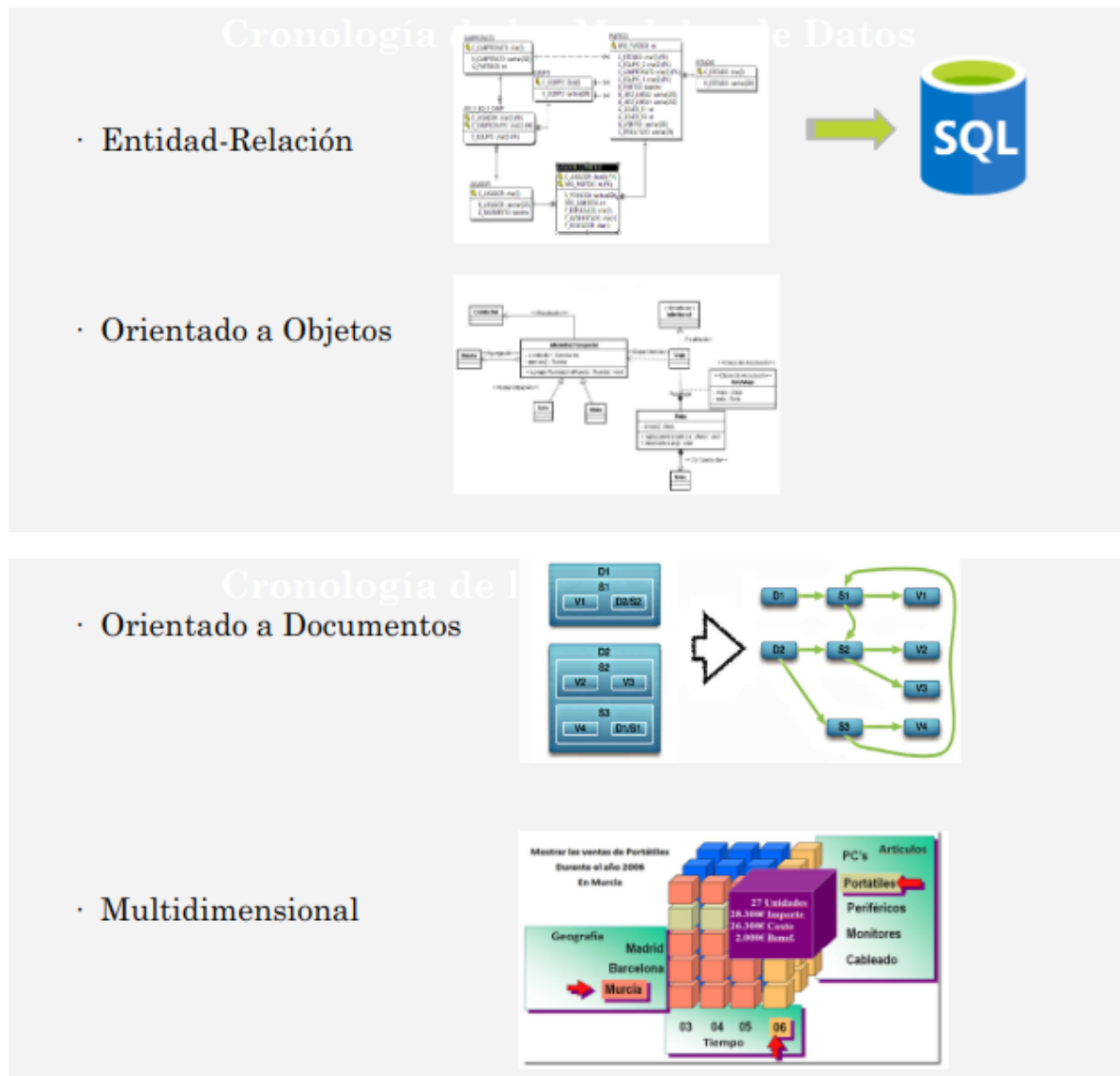


- Jerárquico



- De Red





Modelo Entidad Relación

Se basa en representar al mundo mediante dos objetos básicos: entidades y relaciones.

Entidad: objeto del mundo real que es distinguible de otros objetos. El conjunto de todas las entidades de un mismo tipo se denomina conjunto de entidades.

Relación: asociación entre varias entidades. El conjunto de todas las relaciones del mismo tipo se denomina conjunto de relaciones.

Las entidades se describen mediante un conjunto de atributos.

Por ejemplo: un conjunto de personas que poseen cuentas bancarias.

Entonces Personas y Cuentas Bancarias son entidades.

Atributos de las personas: DNI, Nombre, Dirección y Teléfono.

En este ejemplo, una relación impositora es la que asocia a cada persona con su cuenta bancaria.

Además de entidades y relaciones, el Modelo Entidad Relación representa las restricciones que la Base de Datos debe cumplir.

Este modelo cuenta con una representación gráfica para expresar las estructuras. Esta es llamada Diagrama Entidad Relación.

Diagrama Entidad-Relación: es la representación gráfica de las estructuras lógicas de una Base de Datos.

Los elementos que lo componen son:

- Rectángulo: representan entidades
- Elipses: representan atributos
- Rombos: representan relaciones entre entidades
- Líneas: unen los atributos con las entidades y las entidades con las relaciones.

Modelo Orientado a Objetos

En este modelo la Base de Datos es una colección de objetos.

Un objeto contiene valores almacenados en variables.

Un objeto también contiene fragmentos de código que operan en el objeto llamados métodos.

Los objetos que contienen los mismos tipos de valores y los mismos métodos se agrupan juntos en una entidad llamada clase.

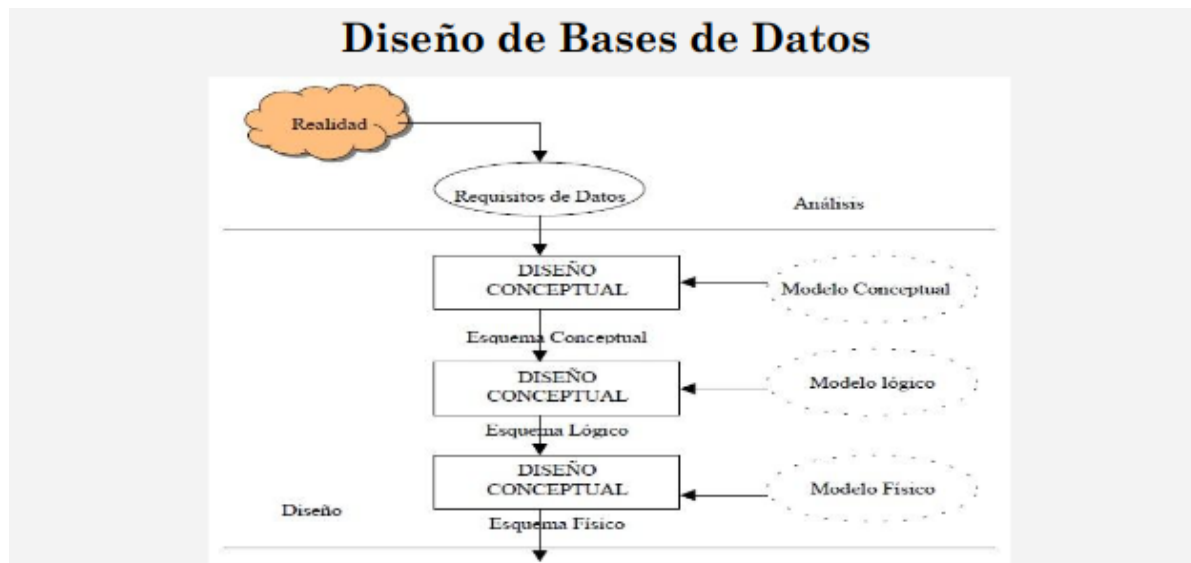
- Los objetos acceden a los datos de otro objeto por medio de un mecanismo llamado pasaje de mensaje.
- La interfaz de llamada de los métodos de un objeto define la parte visible externamente del objeto.
- La parte interna del objeto no es visible externamente. Así se tiene dos niveles de abstracción.
- Cada objeto tiene su propia identidad, la cual es independiente de los datos que contenga.

Modelo lógico basado en registros

- El nombre de modelo lógico basado en registro se debe a que la Base de Datos está estructurada como un conjunto de registros de distintos tipos.
- Cada tipo de registro está definido como un conjunto de atributos o campos.
- Este tipo de modelo es empleado en los niveles físicos y lógicos.
- Se usan tanto para especificar la estructura lógica de la Base de Datos, como para proporcionar una descripción de alto nivel de la implementación.

Existen tres modelos de este tipo que son comúnmente aceptados: el modelo relacional, el modelo jerárquico y el modelo de red.

El modelo relacional es el que ha prevalecido sobre los otros mencionados.



La figura anterior muestra el proceso de diseño de Bases de Datos, donde los requisitos de datos constituyen un componente de los requisitos de un producto y son la entrada al diseño conceptual.

Diseño Conceptual (Nivel conceptual): recibe como entrada la especificación de requerimientos y su resultado es el esquema conceptual de la Base de Datos, que es una descripción de alto nivel de la estructura de la base de datos, independiente del software que se use para manipularla. Este nivel describe cada uno de los objetos o elementos del sistema y sus propiedades, además de las relaciones.

Diseño Lógico (Nivel Lógico): recibe como entrada el esquema conceptual y da como resultado un esquema lógico, que es una descripción de la estructura de la Base de Datos que puede procesar el software de DBMS. En este caso se consideran las restricciones que impone el mecanismo de representación y tratamiento de la información. Los modelos empleados en este nivel son: Relacional, de Redes, entre otros.

Diseño Físico (Nivel Físico): recibe como entrada el esquema lógico y da como resultado un esquema físico, que es una descripción de la implementación de una base de datos en la memoria secundaria, describe las estructuras de almacenamiento y los métodos usados para tener un acceso efectivo a los datos.

Consideraciones sobre diseño de bases de datos relacionales

Es importante conocer exactamente para qué se quiere usar la base de datos, qué datos son los que interesan de los que existen en la realidad y qué información se necesitará extraer.

Una vez que esto esté claro, se definen las Tablas que compondrán la base de datos. Dentro de cada tabla, se piensa qué campos serán necesarios. Es fundamental definir correctamente la base de datos, ya que un mal diseño hará que el sistema sea lento y los resultados no sean los esperados.

Modelo lógico basado en registros

Los SGBD tienen dos lenguajes diferentes, uno para especificar la estructura de la Base de Datos y el otro para expresar las consultas y las actualizaciones.

Lenguajes de definición de datos (DDL)

Permiten especificar los esquemas de las Bases de Datos.

Al compilar las instrucciones de los comandos DDL se obtiene un conjunto de tablas que se almacenan en un archivo especial llamado diccionario de datos o directorio de datos.

Las sentencias DDL se utilizan para crear y modificar la estructura de las tablas así como otros objetos de la base de datos.

- CREATE - para crear objetos en la base de datos.
- ALTER - modifica la estructura de la base de datos.
- DROP - borra objetos de la base de datos.
- TRUNCATE - elimina todos los registros de la tabla, incluyendo todos los espacios asignados a los registros.

Lenguajes de manipulación de datos (DML)

Un lenguaje de manipulación de datos permite a los usuarios acceder o manipular los datos organizados mediante el modelo de datos apropiado logrando:

- Recuperación de información almacenada en la Base de Datos.
- Inserción de información nueva.
- Borrado de información de la Base de Datos.
- Modificación de información almacenada en la Base de Datos.

Como lenguajes DML, podemos mencionar

- Lenguajes procedimentales (requieren de la especificación de los datos y cómo obtenerlos).
- Lenguajes no procedimentales (el usuario especifica los datos sin decir cómo obtenerlos).

Las sentencias de lenguaje de manipulación de datos (DML) son utilizadas para gestionar datos dentro de los schemas.

Ejemplos:

- SELECT - para obtener datos de una base de datos.
- INSERT - para insertar datos en una tabla.
- UPDATE - para modificar datos existentes dentro de una tabla.
- DELETE - para eliminar registros de una tabla.