

Modelo de datos

Javier Fernández Rivera - www.aurea.es

Un modelo: Es una representación de cualquier aspecto o tema extraído del mundo real, en una base de datos esta representación se realizara de forma grafica.

Un modelo de datos: Es un conjunto de conceptos que nos permiten describir los datos, las relaciones que existen entre ellos, la semántica y las restricciones de consistencia.

Existen 3 grupos de modelos de datos

1. **Modelos externos o lógicos basados en objetos:** nos permite representar los datos que necesita cada usuario con las estructuras propias del lenguaje de programación que se vaya a usar.
2. **Modelos globales o lógicos basados en registros:** ayuda a escribir los datos para el conjunto de usuarios.
3. **Modelos físico de datos:** esta orientado a la maquina.

Dentro del modelo global podríamos clasificar 2 tipos:

- Conceptual.
- Convencional.

1. **Los modelos conceptuales:** también denominados de alto nivel facilitan la descripción global del conjunto de información de la empresa con independencia de la maquina que se use.
2. **Los modelos convencionales:** se encuentran soportados por los SGBD, los mas comunes son el jerarquico y el codasyl o de red y el relacional.

Es preciso distinguir entre esquema y ocurrencia.

Esquema: es la descripción de la estructura de la base de datos.

Ocurrencia: son los datos que hay almacenados en el esquema en un determinado momento y que varían.

Una definición de modelo de datos seria: Un conjunto de conceptos, reglas y convenciones que nos permite describir y manipular los datos de un cierto mundo real que deseamos almacenar en la base de datos.

Propiedades de un modelo de datos

Las propiedades de los modelos de datos son de dos tipos:

1. Estáticas
2. Invariantes en el tiempo: Que responden a lo que suele ser las estructuras.
3. Y dinámicas que son las operaciones que se aplican a los datos o valores almacenados en las estructuras, las cuales varían en el tiempo.

Propiedades estáticas

La estática de un modelo de datos esta compuesta por elementos permitidos y elementos no permitidos.

Elementos permitidos: No son los mismos para todos los modelos de datos. Varían especialmente en su terminología, en general suele ser objetos que pueden ser entidades relaciones, registros, asociaciones entre objetos (interrelaciones), propiedades de los objetos (atributos o campos) y dominios que son conjuntos de valores sobre los que se definen las propiedades.

La representación de estos elementos depende del tipo de modelo, así en las jerarquicos y en el modelo en red se representan en forma de grafo (similar o en árbol).

En los modelos relacionales se usan las tablas y en el modelo entidad, interrelación se usan grafos y tablas.

Elementos no permitidos: No todos los valores, cambios de valor o estructuras están permitidas en el mundo real, además de cada modelo de datos también impone por si mismo limitaciones a las estructuras que permite. Se denomina restricciones inherentes (pueden cambiar) a las que son impuestas por el modelo de datos en si y se denominan restricciones de integridad a las que son impuestas por el mundo real. Habitualmente en vez de restricciones se suelen llamar reglas.

Propiedades dinámicas

Los valores que toman los distintos objetos de un esquema en un determinado momento T_i reciben el nombre de ocurrencia de esquema o estado de los datos en el momento BD_i .

En otro momento T_j la ocurrencia sera BD_j .

La aplicación de una operación a una ocurrencia de un esquema trasforma esta en otra ocurrencia.

Una operación tienen dos componentes o pasos:

1. Componente de localización: selección o enfoque, consiste en localizar la ocurrencia de un objeto indicando un camino.
2. Acción, que se realiza sobre la ocurrencia previamente localizada, operaciones de actualización: eliminar, insertar, modificar, ver.

Restricciones de integridad en los modelos de datos

En el mundo real existen ciertas reglas que deben cumplir los elementos existentes.

En la BD deseamos que reflejen lo mas fielmente posible el universo que se representa (si hacemos una BD de un hospital deberá representar elementos del hospital), por lo tanto tenemos que recoger en nuestro sistema de información el universo lo mas fielmente posible para que al desarrollar el esquema de la BD junto con los objetos evocaciones y propiedades que los mismos se cumplan estas reglas llamadas restricciones semánticas o de integridad las cuales pueden ser definidas como condiciones que limitan al conjunto de ocurrencias validas de un esquema, la semántica de los datos que en un principio la controlaba manualmente el usuario.

Hoy en día las restricciones de integridad suelen estar dispersas por la BD en vez de estar dispersas por las diferentes aplicaciones.

Clasificación de los modelos de datos

Durante la década de los 70 y a principios de los 80 unos de los principales temas de discusión era como poder clasificar los modelos de datos. Con la aparición de los niveles de abstracción de la arquitectura ANSI este tema se planifico bastante.

En general nosotros clasificamos los modelos en función de su nivel de abstracción y riqueza semántica así distinguiremos entre conceptual y convencionales. Las características principales de los convencionales son: Están implementados en SGBD comerciales.

Depende del SGBD

Están próximos al ordenador, tienen poca capacidad semántica.

Están mas enfocados a la implantación. Sin embargo los conceptuales no suelen estar implementados en los SGBD, son independientes del SGBD tienen un gran nivel de abstracción, tienen mayor capacidad semántica y están mas enfocados al diseño de alto nivel.

El modelo jerarquico

A este modelo se le llamo en árbol ya que se usa una estructura de árbol invertida con las raíces en la parte superior del diseño para su representación .

Es similar al modelo en red ya que los datos se representan mediante registros llamados nodos y las relaciones se suelen llamar lazos o enlaces.

Sus características son:

Una colección de árboles forman una BD

A los registros se les denomina nodos o segmentos que contienen campos que a su vez se les llama atributos.

Los nodos están organizados en niveles cada nodo tiene campos comunes, vinculados a el.

Padre es un nodo vinculado a otros de nivel inferior.

Gemelos son los nodos al mismo nivel y tienen el mismo padre.

Hojas: son los nodos que no tienen hijos

Altura: Cantidad de niveles

Camino: Son las líneas que unen los nodos no puede haber mas de un camino al mismo sitio o nodo.

Rama: Es un camino que termina en una hoja

Familia: Es el conjunto de padres e hijos, etc.

Momento; Numero total de nodos de un árbol.

Peso: Cantidad de hojas que tiene el árbol.

Modelos de red

Esta compuesto por un árbol que a su vez tiene tablas. Dentro de las características principales podemos decir que los arcos que unen entidades se les llama conjuntos. Los conjuntos están formados por relaciones de muchos a muchos.

El modelo en red es el mas utilizado en el CODASYL aunque hoy en día esta en desuso.