

# HACIA LA INTEGRACIÓN DE BASES DE DATOS RELACIONALES Y BASES DE DATOS DOCUMENTALES.

Gerardo Cerda Neumann \*

## RESUMEN

Al analizar los tipos de datos que son utilizados en cualquier organización, se concluye que existen, básicamente, dos tipos: los estructurados y los poco estructurados. Los primeros son manipulados mediante el uso de las bases de datos relacionales (BDR) y los segundos a través de las bases de datos documentales (BDD). Se propone la creación de una interfaz que integre ambos esquemas. Dicha interfaz es llamada Interfaz Integrada de Bases de Datos (IIBD). Se presenta su base conceptual a partir de la revisión de las características principales de las BDR y BDD. Además se presentan las distintas fases que hay que seguir para construir un esquema conceptual que de sustento a una IIBD. Finalmente, a partir de esta interfaz se presenta un nuevo concepto llamado Base de Datos Integrada (BDI).

## INTRODUCCIÓN

Cuando se busca definir una base de datos que permita solucionar los requerimientos de un usuario resulta evidente que existen de dos tipos: los estructurados y los poco estructurados. Generalmente los primeros son representados mediante tablas y los segundos mediante documentos. Por esta razón surge, en forma natural, la separación entre las bases de datos relacionales, que ven los datos básicamente como tablas, y las bases de datos documentales, que ven los datos como documentos, teniendo éstos poca o ninguna estructura.

Hasta el momento, ambos esquemas se presentan como soluciones separadas, a pesar de que existen productos de software que permiten combinar hasta cierto punto ambas maneras de ver los datos. A raíz de esto, es muy común que, modelar los datos de una situación en particular, se identifiquen solamente los datos con estructura (los que son capturados y manipulados mediante tablas en bases de datos relacionales), siendo los documentales manipulados mediante procesadores de texto o correo electrónico. Esta situación, muy común en las organizaciones, presenta la gran desventaja de que resulta muy difícil encontrar las relaciones entre ambos tipos de datos en forma rápida y oportuna.

\* Ingeniero Ejecución en Informática, U. de Santiago. Profesor Escuela de Ingeniería, UCINF.

La realidad es que existe una estrecha unión entre ambos tipos de datos, la que se traduce en que unos son creados como consecuencia de los otros. Así por ejemplo, los compromisos de pago de un alumno universitario que se registran en una tabla, son creados a partir de la firma de un documento llamado matrícula anual para una carrera profesional. También ocurre a la inversa, ya que el análisis estadístico hecho a los cumplimientos de estos compromisos puede dar origen a un documento que concluya que una carrera no resulta rentable.

Como se puede notar, ambos datos son absolutamente complementarios, a pesar de esta disociación que existe normalmente en los sistemas informáticos de una organización.

Por este motivo, se analizarán las principales características que poseen, tanto las bases de datos relacionales (BDR), como las bases de datos documentales (BDD), para incluir las más relevantes en la propuesta de una interfaz integrada de bases de datos (IIBD). Para mayor claridad, las características más relevantes serán destacadas mediante un número encerrado entre corchetes [1].

## 1. BASES DE DATOS RELACIONALES

### 1.1. Introducción.

Sin duda que las BDR son muy usadas actualmente (se calcula que alrededor del 90% de las bases de datos utilizadas son relacionales). Esto se debe a que permiten solucionar muchas situaciones en las que se necesita manejar datos estructurados. A modo de ejemplo se puede citar: las cuentas corrientes de un banco, las notas de los alumnos en un colegio y las ventas diarias en un negocio.

Desde que fue propuesto este modelo por Codd en 1970, se transformó en un hito dentro de la breve historia de las bases de datos. El objetivo principal de este modelo fue independizar al usuario de las características físicas de los dispositivos de almacenamiento (De Miguel, 2000). Además proporcionan una gran sencillez y uniformidad, una sólida fundamentación teórica (el álgebra y cálculo relacional) y la independencia de la interfaz del usuario.

Rápidamente se crearon métodos que permitieron desarrollar grandes bases de datos de manera rápida y confiable.

A continuación se presentarán las principales características que poseen las BDR, así como su utilización.

## 1.2. Principales características de las BDR

Al estar sustentadas en el álgebra y el cálculo relacional, las BDR se componen básicamente de: entidades, atributos, dominios y relaciones. La combinación de estos elementos permite construir soluciones computacionales que satisfagan los requerimientos (necesidades) de los usuarios. Es por esto que:

- El diseño de soluciones que incluyen bases de datos relacionales es el proceso por el cual se determina (Batini, 1994):
  - La organización de la base de datos, que incluye la estructura (definida a partir de la abstracción de los datos) y el contenido.
  - Las aplicaciones por desarrollar que permiten manipular (insertar, borrar, modificar y consultar) los datos.
- El diseño de la base de datos se realiza después de identificar claramente los requerimientos del usuario (Elmasri, 1997). Por este motivo el usuario debe estar en el centro del análisis y el diseño [1].
- El diseño se realiza en tres fases:
  - Diseño conceptual: produce una representación abstracta de alto nivel de la realidad del problema. Se identifican las entidades (con sus atributos) que permiten solucionar el problema definido en los requerimientos [2].
  - Diseño lógico: convierte esta representación en especificaciones que pueden implantarse en un computador y ser procesados por él [3].
  - Diseño físico: determina las estructuras de almacenamiento físico y los métodos de consulta requeridos para un acceso eficaz a los contenidos [4].
- Se define entidad como clases de objetos de la realidad. Es una “cosa” del mundo real con existencia independiente. Puede tener existencia física o conceptual [5].
- El modelo entidad-relación describe los datos como: entidades, vínculos y atributos (el dominio es el conjunto de valores posibles para un atributo) [6].

- Las entidades “débiles” son aquellas que existen solamente porque existen otras y además no tienen atributos clave propios [7].
- Se puede realizar un diseño de vistas (la parte de la base de datos que le corresponde usar a un usuario específico) a partir de formularios ya existentes. Para esto se realiza:
  - Un análisis de requerimientos [8] donde:
    - Se distinguen las partes extensivas, intensivas y descriptiva del formulario.
    - Se seleccionan las áreas y subáreas.
  - Diseño inicial que consiste en construir un esquema de “armazón global”.

## 2. BASES DE DATOS DOCUMENTALES

### 2.1. Introducción.

Por documento se entiende (RAE, 1992): “escrito en que constan datos fidedignos o susceptibles de ser empleados como tales para probar algo”. También se define como “registro de información estructurado para consumo humano” (Sprague, 1994).

De estas definiciones se pueden extraer tres conclusiones:

- Un documento lleva como elemento básico texto escrito (aunque puede incluir además imágenes, enlaces de hipertexto y tablas) [9].
- Los datos que posee deben ser fidedignos, es decir deben estar sustentados por alguna fuente de datos que sea correcta.
- Se crean para ser usados por personas que los necesitan. Por esta razón sólo van a existir cuando se identifican primero los requerimientos del usuario.

Las BDD nacen para realizar la gestión de los documentos al interior de las organizaciones.

### 2.2. Características de los documentos y de las BDD.

Los documentos poseen las siguientes características (Elmasri, 1997):

- Una gran riqueza semántica, es decir, poseen muchas formas de ser interpretados [10].
- Los documentos usados en las oficinas poseen poca estructuración. Usualmente, los documentos usados en las oficinas poseen un encabezado que explicita quién lo envía, quién es el destinatario, en qué fecha se generó y qué tema trata. Este encabezado más estructurado, a veces, puede ser la portada de un informe o bien las primeras líneas de un memorándum [11].
- Existen “agrupaciones de información estereotípica”, como [12]:
  - Cartas de negocio.
  - Informes trimestrales.
- Un documento debe incorporar el factor tiempo, es decir debe permitir:
  - Capturar el tiempo total de “recorrido” (uso consecutivo del documento por parte de todas las personas interesadas en él) [13]
  - La duración de las actividades que define, junto a sus calendarios y planes.
  - El tiempo permitido para responder un documento que ha sido enviado.
- Además, un documento debe permitir:
  - Coordinar actividades que se realizan al interior de la organización [14].
  - Ser filtrado para que sólo las personas autorizadas puedan verlo y/o modificarlo [15].
  - Generar recordatorios a los integrantes de la organización para que lo revisen y/o actualicen en los lugares que le corresponde [16].
- En un documento el orden físico de los datos es muy relevante, al contrario de lo que ocurre en una base de datos relacional (Sengupta, 1996) [17].
- Un documento es un sistema de ideas. Algunas ideas son consolidaciones simbólicas de tablas valorizadas. Por ejemplo: “el alza del dólar” es una idea documental ya que puede estar expresada en un documento. Esta idea puede estar respaldada por una tabla (información estructurada) que indica la variación de la moneda versus el tiempo transcurrido en días. Ningún valor en la tabla por sí solo respalda esta idea sino toda la tabla [18]. No todas las ideas son respaldadas por tablas o, genéricamente, por información estructurada, como ideas políticas o filosóficas (Covacevich, 2001).

### 3. INTERFAZ INTEGRADA DE BASES DE DATOS (IIBD).

#### 3.1. Introducción.

La IIBD busca sacar provecho de las mejores características tanto de las BDR como de las BDD.

¿Por qué se llama integrada? Para clarificar esto se presenta la definición del concepto integrar: “construir las partes de un todo, completar un todo con las partes que faltan” (Pequeño Larousse, 1994). Justamente éste es el objetivo de la IIBD: poder manipular todos los datos existentes en una organización, tanto los estructurados como los poco o no estructurados, a través de la integración de las bases de datos relacionales con las documentales. Para esto se realiza una agregación entre las BDR y las BDD que permite definir un concepto llamado Base de Datos Integrada o BDI (Cerdea, 2001). Sin embargo, como aún no existe un software que la implemente, se ha preferido referirse solamente a la interfaz.

Para poder definir esta interfaz deben existir datos en común entre ambas bases de datos. Estos datos son llamados “datos de enlace” ya que permiten asociar los datos entre ambos esquemas.

#### 3.2. Características de la IIBD.

A partir del análisis presentado en los apartados anteriores, es posible identificar las siguientes características que debe cumplir la IIBD:

- [1] Una IIBD se construye a partir de requerimientos específicos de los usuarios.
- [2] Una IIBD posee un esquema conceptual que indica sus entidades, atributos y datos de enlace.
- [3] Una IIBD posee un esquema lógico en el que se especifica cuáles de las entidades serán tablas, cuáles documentos y cuáles serán los datos de enlace que se usarán. Este esquema lógico es independiente de un software de base de datos específico.
- [4] Una IIBD posee un esquema físico donde se especifican los componentes con las características físicas y restricciones que les imponen las bases de datos seleccionadas.
- [5] Una IIBD debe poder representar cualquier entidad que sea relevante para solucionar los requerimientos.
- [6] El modelo entidad relación que se construye para especificar el esquema conceptual, utiliza tanto los símbolos como las notaciones estándar.

- [7] Una IIBD debe poder representar tanto entidades fuertes como débiles.
- [8] Una IIBD debe poder solucionar todos los requerimientos definidos.
- [9] Un elemento que se incluya en una IIBD puede ser totalmente estructurado (como una tabla), medianamente estructurado (como un memorándum) o poco estructurado (como una carta personal), pudiendo incluso ser una imagen e incluir enlaces de hipertexto.
- [10] Una IIBD puede contener información de gran riqueza semántica que puede ser utilizada por el usuario para apoyar su gestión dentro de la organización.
- [11] Una IIBD debe permitir incluir cualquier documento utilizado por la organización respetando su formato original o bien traspasándolo a un formato totalmente relacional o documental. Esta decisión se hace únicamente con el objetivo de mejorar la gestión de la información.
- [12] Los elementos que componen una IIBD deben indicar claramente quiénes son los dueños/creadores si existen.
- [13] De ser necesario, los elementos incluidos en una IIBD deben permitir registrar el tiempo de uso de él.
- [14] De ser necesario, una IIBD debe permitir coordinar actividades de uso de la información al interior de la organización.
- [15] Los elementos que son referenciados a través una IIBD deben ser accedidos y/o modificados solamente por los usuarios autorizados.
- [16] Una IIBD debe permitir generar recordatorios a los usuarios para facilitar el uso de la información que contiene.
- [17] Una IIBD debe poder registrar el orden físico de los documentos que contiene.
- [18] Al incluir documentos que son un sistema de ideas, una IIBD debe permitir hacer referencias y enlaces entre los datos con poca estructura y los datos estructurados. Estos serán en parte los datos de enlace.

### 3.3. Fases para la construcción de una IIBD.

Una vez definidas las características que debe cumplir una IIBD para que resulte útil, se presentan las fases necesarias para poder construirla.

Para construir una IIBD se proponen seis fases, a saber:

## **Fase 1: Análisis de requerimientos.**

Objetivo: definir claramente qué es lo que necesita el usuario.

En primer lugar, se listan todas las necesidades del usuario.

Se identifican los documentos que utiliza el usuario. Estos documentos pueden ser en papel o cualquier tipo de soporte.

## **Fase 2: Identificación de entidades.**

Objetivo: identificar todas las entidades reconocibles en la situación analizada.

Se define entidad como: “cualquier cosa que existe en la situación analizada y que es relevante” o bien como “cualquier objeto (real o abstracto) que existe en la realidad analizada y acerca del cual queremos almacenar información en la base de datos” (De Miguel, 2000). Una característica de una entidad es que deben existir varios ejemplares de ella. De lo contrario no se representa.

Para identificar las entidades se deben responder las siguientes preguntas:

1. ¿Qué datos se necesitan para solucionar los requerimientos identificados en la fase 1?
2. ¿Esos datos están disponibles?
3. Si lo están, ¿en qué formato? Se listan todos los documentos y tablas que contienen estos datos.
4. ¿Qué datos se necesitan y no se poseen actualmente?
5. ¿Cómo se identifica en forma única una entidad? Se revisan los documentos existentes que poseen datos relevantes y se identifican las palabras más importantes dentro de ellos. Cada una de estas palabras son candidatas a transformarse en una entidad. Posteriormente, se reúnen las palabras que representan el mismo significado y se elige un nombre que las represente a todas (puede ser una de ellas). Se consideran estas palabras como entidades. Se contesta la pregunta previa para cada una de estas palabras. Esto permite identificar las claves principales de las entidades (Barker, 1994).

6. ¿Cómo se estructuran los documentos? Al revisar cada documento se identifican las partes, que son:

- Información del encabezamiento.
- El detalle.
- Detalle de nivel inferior.
- Información de referencia.

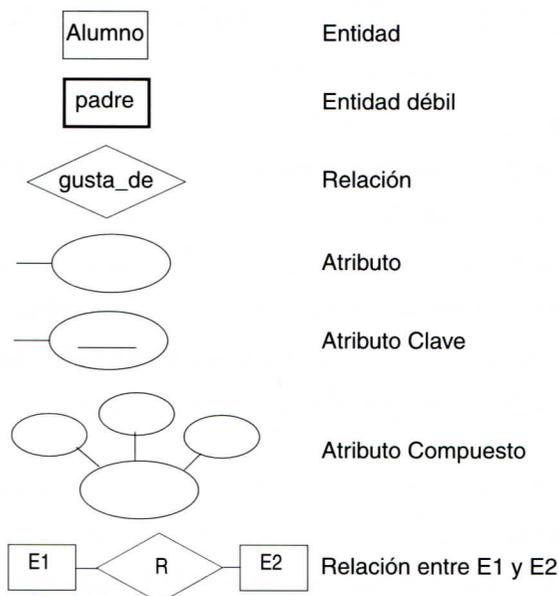
Cada una de estas partes permite identificar entidades relevantes. Además, se pueden ya definir las interrelaciones entre las entidades. Probablemente va a existir una interrelación una a muchos entre las entidades del encabezamiento y las de detalle.

7. ¿Qué información necesita conocer o mantener acerca de una entidad? Esto permite identificar los atributos de las entidades.

### Fase 3: Esquema conceptual.

**Objetivo:** construir un esquema conceptual usando el modelo entidad relación (Chen, 1976).

Este esquema conceptual se compone de las entidades, atributos e interrelaciones identificadas en la Fase 2. Este esquema conceptual es construido usando los siguientes símbolos:



### Fase 4: Esquema lógico.

**Objetivo:** construir un esquema lógico que especifique cuáles de las entidades van a ser tablas y cuáles documentos.

Para esto se deben responder las siguientes preguntas:

¿La entidad nace de un documento que resulta necesario para resolver los requerimientos? Si la respuesta es sí entonces es candidata a ser representado por un documento.

¿La entidad posee una clara estructura? Si la respuesta es sí entonces es candidata a ser representada por una tabla. Si la respuesta es no entonces es candidata a ser representada por un documento.

¿Las entidades candidatas a ser representadas mediante documentos poseen un dueño, es decir alguna persona que genera los diferentes ejemplares? Si la respuesta es sí entonces se define este dueño como atributo y se agrega además un título descriptivo. Si corresponde se agregan los atributos que indiquen:

- El o los destinatarios (que pueden ser o no parte del dominio del atributo dueño).
- Fecha de creación del documento.
- Fecha o fechas de actualización del documento.
- Tabla de contenido del documento.

¿Cuáles son los atributos que comparten las entidades candidatas a ser representadas por tablas con las candidatas a ser representadas por documentos? Estos atributos se definen como “**Datos de enlace**” y posteriormente permitirán que tanto las tablas como los documentos se relacionen en la IIBD.

### **Fase 5: Esquema físico.**

Objetivo: definir los atributos físicos de las entidades y atributos (Chen, 1976).

Este esquema describe todas las entidades y atributos identificadas en la Fase 3. Se utiliza las características físicas propias tanto de la Base de Datos Relacional como Documental seleccionada. También se decide si algunos de los documentos que se utilizarán en la IIBD se mantendrán en su formato original (por ejemplo de procesador de texto) y sólo serán descritos y referenciados mediante una tabla relacional llamada “Documentos externos” que posee los siguientes atributos:

- Identificación del documento (este atributo normalmente va a ser el nombre del archivo externo). Es parte de la clave principal.
- Ubicación física del documento (este atributo describe la carpeta donde se ubica el documento referenciado). Es parte de la clave principal.
- Fecha última modificación. Es parte de la clave principal.
- Descripción del documento.
- Dueño (autor del documento o la persona que solicitó que se creara).

### **Fase 6: Construcción de las BDR, BDD y IIBD.**

Objetivo: construir físicamente tanto las bases de datos relacionales y ocutmentales como la interfaz. En esta fase se construye físicamente cada base de datos. Posteriormente se realiza la programación que permita asociar ambas bases de datos a través de los datos de enlace. No se entrará en mayor detalle al respecto, dado que el objetivo de este documento es presentar el aspecto más conceptual de la IIBD.

#### 4. CONCLUSIONES

Al analizar cualquier situación en la que se necesita manejar información, es fácil identificar la presencia de datos que son estructurados, otros que son poco estructurados y otros que no lo son. Tradicionalmente se ha resuelto el manejo de los datos estructurados mediante el uso de las bases de datos relacionales y los no estructurados mediante el uso de las bases de datos documentales. Sin embargo, los datos que poseen poca estructura, generalmente, son manejados a través de archivos en formato de procesador de texto o planillas de cálculo, siendo enviados a través de correo electrónico. Esto da lugar a numerosas versiones de documentos y a problemas para compartirlos.

Por estas razones, se ha propuesto el desarrollo conceptual de una base de datos integrada que permita sacar el mayor provecho a ambos esquemas. Como es una propuesta conceptual, se da libertad para usar el RDBMS y el software documental que se quiera. La única condición es que ambos softwares puedan relacionarse. Esta exigencia impone una dificultad fácil de solucionar mediante programación. Los beneficios, sin embargo, son mucho mayores:

- Integración de todos los datos en el mismo esquema conceptual.
- Posibilidad de coordinar a todos los usuarios de los datos fácilmente.
- Facilidades de búsqueda y recuperación de documentos completos que tengan referencias precisas a tablas.

Se espera que esta propuesta conceptual permita, más adelante, construir una Base de Datos Integral que permita tener directamente y mediante un solo software todos estos datos que actualmente están dispersos por la organización. Las BDI deberían por tanto realizar un aporte significativo para la gestión de información al interior de cualquier organización.

## BIBLIOGRAFÍA

**BARKER, Richard (1994).** El modelo entidad - relación, Addison Wesley / Díaz de Santos, 241 p.

BATINI, Carlo; CERI, Stefano y NAVATHE, Shamkant B. (1994). **Diseño conceptual de bases de datos: un enfoque de entidades - interrelaciones.** Addison Wesley / Díaz de Santos, 546 p.

CERDA NEUMANN, Gerardo (2001). "Bases de Datos Integradas", **Revista Akádemeia**, Agosto 2001.

CHEN, P.P.S. (1976). "The entity - relationship model. Toward a unified view of data", **ACM Transactions on Database Systems**, v.1, N° 1, March, 1976.

COVACEVICH VIEIRA, Alejandro (2001), **Los documentos como sistemas de ideas**, Santiago, UCINF. (Documento de uso interno)..

DE MIGUEL, Adoración; PIATTINI, Mario y MARCOS, Esperanza (2000) **Diseño de bases de datos relacionales**, Alfaomega - RA-MA, 549 p..

ELMASRI, Ramez y NAVATHE, Shamkant B. (1997). **Sistemas de bases de datos: Conceptos fundamentales.** Addison Wesley Iberoamericana, 887 p..

EDICIONES LAROUSSE (1994). **Pequeño Larousse en color.** Buenos Aires, 1564 p.

REAL ACADEMIA DE LA LENGUA ESPAÑOLA, (1992). **Diccionario de la lengua española**, 21° ed.

SENGUPTA, Arijit, (1996). "Structured Document Databases": "<http://citeseer.nj.nec.com/sengupta95design.html>", 8 p.

SPRAGUE, Ralph H., Jr. (1994), "Electronic document management: challenges and opportunities for information systems managers": "[http://www.documentmanagement.org.uk/articles/Challenge\\_Opport.htm](http://www.documentmanagement.org.uk/articles/Challenge_Opport.htm)", 37p.