

1.4 Sistemas de bases de datos frente a los sistemas de archivos

Si analizamos la situación (especialmente la situación pasada) de algunos sistemas de información, podemos encontrar sistemas en los que existe una proliferación de ficheros, específicos cada uno de ellos de una determinada aplicación (véase figura 1.8). Los datos se recogen varias veces y se encuentran repetidos en los distintos archivos. Esta redundancia, además de malgastar recursos, origina a menudo divergencias en los resultados.

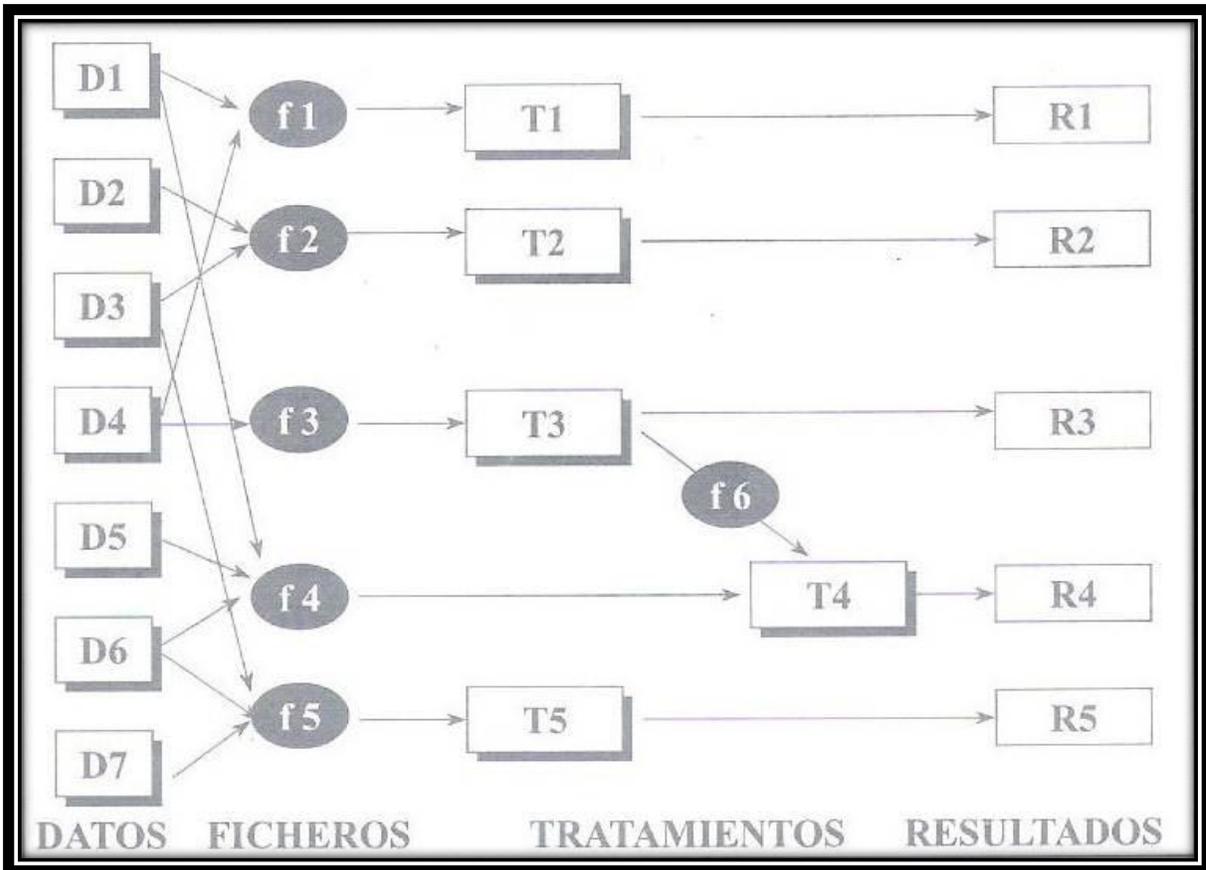


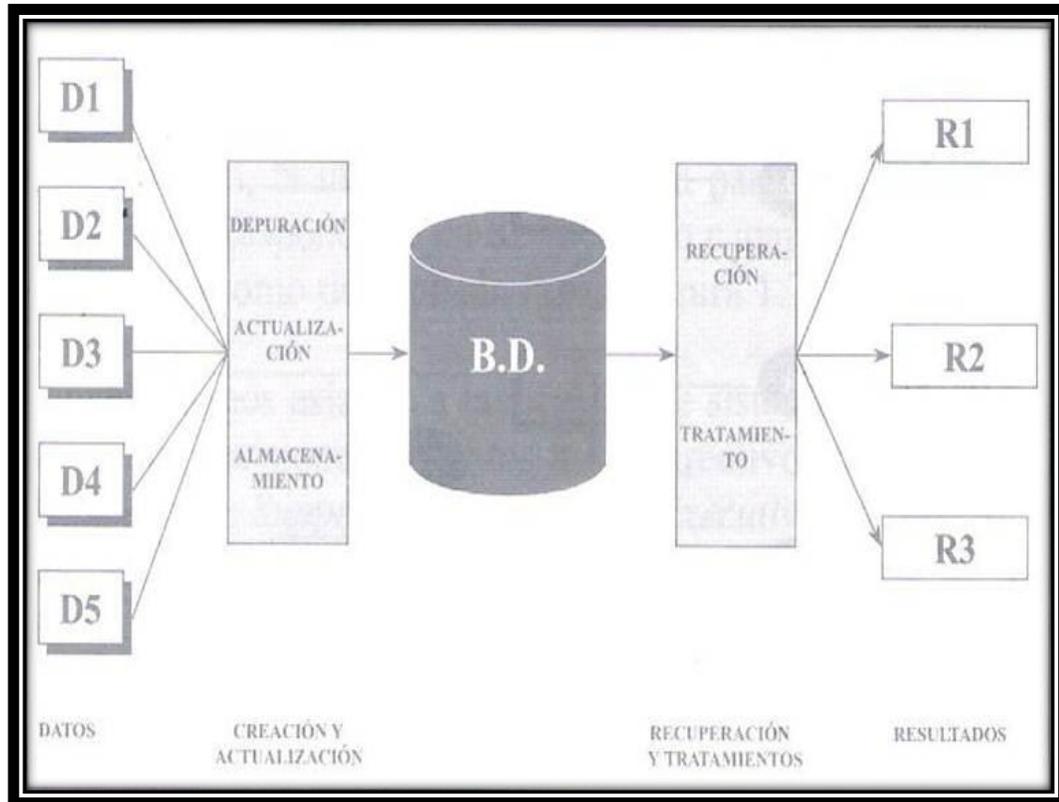
Figura 1.8. Organización clásica: Sistemas orientadas a procesos.

Los sistemas informáticos de este tipo han sido llamados a veces sistemas orientados hacia el proceso, debido a que en ellos se pone el énfasis en los tratamientos que reciben los datos, los cuales se almacenan en ficheros diseñados para una determinada aplicación. Las aplicaciones se analizan e implantan con entera independencia una de otras, y los datos no se suelen transferir entre ellas, sino que se duplican siempre que los correspondientes trabajos los necesitan.

Este planteamiento produce, además de una ocupación inútil de memoria secundaria, un aumento de los tiempos de proceso, al repetirse los mismos controles y operaciones en los distintos ficheros. Pero más graves todavía son las inconsistencias que a menudo se presentan en estos sistemas, debido a que la actualización de los mismos datos, cuando se encuentran en más de un fichero, no se suele realizar de forma simultánea en todos ellos.

Por otra parte, la dependencia de los datos respecto al soporte físico y a los programas da lugar a una falta de flexibilidad y de adaptabilidad frente a los cambios que repercute muy negativamente en el rendimiento de conjunto del sistema informático.

De este análisis se deduce claramente la necesidad de una gestión más racional del conjunto de datos surgiendo así un nuevo enfoque que se apoya sobre una base de datos, en la cual los datos son recogidos y almacenados una sola vez, con independencia de los tratamientos (véase figura 1.9)



Las bases de datos, surgidas como respuesta al nuevo planteamiento de los sistemas orientados hacia los datos para mejorar la calidad de las prestaciones de los sistemas informáticos y aumentar su rendimiento, presentan una multitud de ventajas frente a los sistemas clásicos de ficheros.

Debemos, sin embargo, señalar, para que nadie se sienta defraudado por el fracaso de algún proyecto de este tipo, que las bases de datos no son la panacea universal que soluciona todos los problemas que la información plantea a las empresas. Las bases de datos son un instrumento, que supone un distinto enfoque en la gestión de los datos, y su éxito o su fracaso estará condicionado por el uso que de ella sepamos hacer, no solo los técnicos, sino también los directivos.

Las ventajas de los sistemas de bases de datos son, entre otras, las siguientes:

- A) Independencia de los datos respecto a los tratamientos y viceversa.-** La mutua independencia de datos y tratamientos lleva a que un cambio de estos últimos no imponga un nuevo diseño de la base de datos. Por otra parte, la inclusión de nuevas informaciones, desaparición de otras, cambios en la estructura física o en los caminos de acceso, etc., no deben obligar a alterar los programas. Esta independencia de los tratamientos frente a la estructura de la base de datos, supone una considerable ventaja, al evitar el importante esfuerzo que origina la reprogramación de las aplicaciones cuando se producen cambios en los datos.
- B) Coherencia de los resultados.-** Debido a que la información de la base de datos se recoge y almacena una sola vez, en los tratamientos se utilizan los mismos datos, por lo que los resultados de todos ellos son coherentes y perfectamente comparables. Además, al no existir (o al menos disminuir en gran medida) la redundancia en los datos, desaparece el problema que se presentaba en el enfoque clásico de que el cambio de un dato obligaba a actualizar una serie de ficheros. De esta forma se elimina también el inconveniente de las divergencias en los resultados debidas a actualizaciones no simultáneas en todos los ficheros.
- C) Mejor disponibilidad de los datos para el conjunto de los usuarios.-** Cuando se aplica la metodología de bases de datos, cada usuario ya no es propietario de los datos, puesto que estos se comparten entre el conjunto de aplicaciones, existiendo una mejor disponibilidad de los datos para todos los que tienen necesidad de ellos, siempre que estén autorizados para su acceso. Hay también una mayor transparencia respecto a la información existente, ya que todos los datos que se encuentran en la base se deben relacionar en un catálogo o diccionario que puede ser ampliamente difundido y accedido por medios informáticos.
- D) Mayor valor informativo.-** Puesto que la base de datos ha de ser reflejo del mundo real, en ella se recogen las interrelaciones entre los datos, por lo que el valor informativo del conjunto es superior a la suma del valor informativo de los elementos individuales que lo constituyen; es decir, actúa el efecto de sinergia.
- E) Mejor y más normalizada documentación de la información, la cual está integrada con los datos.-** En el enfoque clásico los datos se encuentran separados de su contenido semántico; los primeros se almacenan en ficheros y su descripción se hace mediante un lenguaje de programación que se encuentra en los programas. La documentación de los datos, realizada por el analista o programador, es en general insuficiente, y a veces incluso inexistente. Además, por lo común, la estandarización brilla por su ausencia. Este problema se atenúa en gran medida en las bases de datos, ya que en la misma base se incluyen no solo los datos, sino también la semántica de los mismos.
- F) Mayor eficiencia en la recogida, validación e introducción de los datos en el sistema.-** Al no existir apenas redundancias, los datos se recogen y validan una sola vez, aumentando así el rendimiento de todo el proceso previo al almacenamiento.
- G) Reducción del espacio de almacenamiento.-** La desaparición (o disminución) de las redundancias, así como la aplicación de técnicas de compactación, lleva en los sistemas de

bases de datos a una menor ocupación de almacenamiento secundario – disco magnético
-. Se ha de tener presente, sin embargo, que los elementos del sistema (diccionario, referencias, punteros, ficheros invertidos, etc.) ocupan bastante espacio.

VENTAJAS DE LAS BASES DE DATOS

REFERIDAS A:

A) LOS DATOS:

- Independencia de estos respecto de los tratamientos y viceversa.
- Mejor disponibilidad de los mismos.
- Mayor eficiencia en la recogida, codificación y entrada en el sistema.

B) LOS RESULTADOS:

- Mayor coherencia.
- Mayor valor informativo.
- Mejor y más normalizada documentación.

C) LOS USUARIOS:

- Acceso más rápido y sencillo de los usuarios finales.
- Más facilidades para compartir los datos por el conjunto de los usuarios.
- Mayor flexibilidad para atender a demandas cambiantes.

Como hemos señalado anteriormente, las bases de datos no solo presentan ventajas, sino que también tienen posibles inconvenientes que es necesario valorar antes de tomar una decisión relativa a un cambio en la orientación del SI. Entre estos inconvenientes es preciso destacar:

- A) **Instalación costosa.**- La implantación de un sistema de bases de datos puede llevar consigo un coste elevado, tanto en equipo físico (nuevas instalaciones o ampliaciones), como en el lógico (sistemas operativos, programas, compiladores, etc. Necesarios para su uso), además del mismo coste de adquisición y mantenimiento del SGBD.
- B) **Personal especializado.**- Los conocimientos, que resultan imprescindibles para una utilización correcta y eficaz y sobre todo para el diseño y administración de las bases de datos, implican una necesidad de personal especializado.
- C) **Implantación larga y difícil.**- Debido a las causas apuntadas anteriormente, la implantación de una base de datos puede convertirse en una tarea larga y laboriosa. Las dificultades que van apareciendo a los largo de su desarrollo llevan en general a que se superen ampliamente los plazos inicialmente previstos.

- D) **Falta de rentabilidad a corto plazo.**- La implantación de un sistema de bases de datos, tanto por su coste en personal y en equipos como por el tiempo que tarda en estar operativo, no resulta rentable a corto plazo, sino a medio o, incluso, a largo plazo.
- E) **Escasa estandarización.**- Un problema muy importante que se pone en manifiesto en el momento de la creación de una base de datos, es la falta de estandarización que facilite a los usuarios el manejo de los sistemas de bases de datos. Empieza, sin embargo, a observarse ya una preocupación por este tema, y van apareciendo estándares, tanto de iure como de facto, sobre todo el campo de las bases de datos relacionales (SQL92), y actualmente también en las bases de datos orientadas a objetos (ODMG-93 Y SQL3).
- F) **Desfase entre teoría y práctica.**- Al existir un considerable avance de la teoría en relación con la práctica, en muchas ocasiones los usuarios, especialmente los directivos, se engañan respecto a las prestaciones reales que pueden proporcionarles los SGBD actuales, creyendo que constituyen ya una realidad de ciertos aspectos que todavía son solo teóricos.

1.5 Los distintos niveles de abstracción de una base de datos

La estructura lógica de usuario o esquema externo es la visión que tiene de la base de datos cada usuario en particular; la estructura lógica global (también denominada esquema conceptual) responde al enfoque del conjunto de la empresa y la estructura física (o esquema interno) es la forma en que se organizan los datos en el almacenamiento físico. La estructuración de una base de datos en estos tres niveles de abstracción tiene como principal objetivo conseguir la independencia entre datos y aplicaciones.

ESTRUCTURA LOGICA DE USUARIO: ESQUEMA EXTERNO

Debido a que un esquema externo es la visión que de la base de datos tiene un usuario en particular, en él deberán encontrarse reflejados sólo aquellos datos e interrelaciones que necesite el correspondiente usuario. También habrán de especificarse las restricciones de uso, como puede ser el derecho a insertar o borrar determinados datos o el acceso a los mismos, etc.

Aunque esto no sea lo más conveniente, ya que indica una dependencia físico-lógica, pueden que aparezcan en este nivel los caminos de acceso a los datos, hecho que dependerá en gran medida del modelo de los datos en el que se apoya el correspondiente SGBD; en el modelo relacional los caminos de acceso sólo se encuentran en el nivel, no siendo nunca visibles por los usuarios.

Habrán tantos esquemas externos como exijan las diferentes aplicaciones. Un mismo esquema externo podrá ser utilizado por varias aplicaciones.

ESTRUCTURA LOGICA GLOBAL: ESQUEMA CONCEPTUAL

En el esquema conceptual, por ser la visión global de los datos, deberá incluirse la descripción de todos los datos e interrelaciones entre estos, así como las restricciones de integridad y de confidencialidad. La estabilidad de estos conceptos disminuye en el orden en el que los hemos citado. Así, las restricciones de confidencialidad serán menos estables que las de integridad, y estas, a su vez, serán menos estables que las interrelaciones o que los datos. Por esta razón, algunos autores proponen que este esquema se divida en varios, uno para cada concepto, de modo que, por ejemplo, un cambio en las restricciones no lleve consigo una nueva definición de todo el esquema.

ESTRUCTURA FISICA: ESQUEMA INTERNO

Aunque el contenido del esquema interno depende mucho de cada SGBD, podemos distinguir tres clases de aspectos que deben especificarse en él.

- Estrategia de almacenamiento.- En este apartado se incluye la asignación de espacios de almacenamiento para el conjunto de datos. También deberá indicarse la estrategia de emplazamiento de los datos que ha sido utilizada para optimizar tiempos de respuestas y espacio de memoria secundaria; por último deberán aparecer aspectos como el tratamiento de los desbordamientos, etc.
- Caminos de acceso.- Incluimos en los caminos de acceso la especificación de claves, así como la de índices o punteros.
- Miscelánea.- Además de los aspectos citados, habría que incluir, en el esquema interno, otros varios, como técnicas de compresión de datos, de criptografiado, la correspondencia entre esquema interno y esquema conceptual, técnicas de ajuste o afinamiento (tuning), optimización, etc.

El administrador de la base de datos habrá de especificar:

- Dispositivos de memoria: Tamaño de la página, número de páginas asignadas a cada área de almacenamiento, tamaño de las áreas de entrada/salida (buffers), etc.
- Correspondencia entre esquemas (mapping): Por omisión, se suele suponer que existe una correspondencia uno a uno entre los registros del esquema conceptual y los registros almacenados; en caso contrario, el administrador debe indicar la relación existente entre ellos.
- Organizaciones físicas: Para mejorar la recuperación y los tiempos de acceso, el sistema debe dar facilidades para que el administrador defina el tipo de organización (dispersión –hashing-, agrupamientos, índices, etc.) que considere más adecuada a fin de lograr la máxima eficiencia; dependiendo del SGBD podrá definir también definir punteros entre registros, privilegiando así determinados caminos de acceso.

Controles de acceso: Permite definir reglas para proteger la confidencialidad de los datos.