

1. Bases de datos e sistemas de almacenamento da información

1.1 Sistemas lóxicos de almacenamento da información

1.1.1 Sistemas de información

Introdución

Hoxe en día, os Sistemas de Información xogan un papel cada vez mais importante nas modernas organizacións empresariais, ata o punto de condicionar o seu éxito ou fracaso nun entorno económico e social tan dinámico e turbulento como o que caracteriza ó mundo actual.

Novos fenómenos como a globalización ou o tránsito cara unha economía mais baseada no coñecemento induciron importantes cambios nas organizacións empresariais. Neste novo contexto, os Sistemas e as Tecnoloxías da Información e Comunicación (TICs) convertéronse nun elemento esencial como motor do cambio e fonte de vantaxes competitivas.

A información e as novas tecnoloxías que a soportan aparecen como un novo factor produtivo que se suma ós factores tradicionais, traballo e capital, diferenciándose destes últimos polo seu carácter intanxible, dando lugar a un novo escenario no que a información e o coñecemento son básicos para as empresas. Os Sistemas e Tecnoloxías da Información convértense, polo tanto, nunha peza fundamental para propiciar este cambio e este novo escenario ó que nos referimos.

Dentro dunha organización, o Sistema de Información actúa coma o "sistema nervioso" xa que este é o que se encarga de facer chegar a tempo a información que necesitan os distintos elementos da organización empresarial (departamentos, áreas funcionais, equipos de traballo, delegacións, etc.), permitindo desta forma unha actuación conxunta e coordinada, áxil e orientada a conseguir os obxectivos da organización.

Os Sistemas de Información adquiriron unha dimensión estratéxica nas empresas do novo milenio e deixaron de ser considerados unha simple ferramenta para automatizar procesos operativos, para convertérense nunha peza clave a ter en conta á hora de formular a estratexia empresarial para levar a cabo a súa implantación e para realizar o control da xestión.

Clasificación dos subsistemas dunha organización

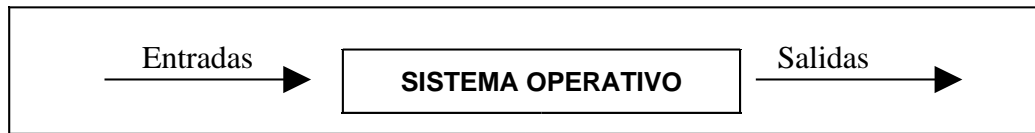
En xeral, podemos definir un **sistema** como un “conxunto de elementos que interactúan entre si, orientados á consecución dun obxectivo común”.

Calquera empresa ou organización é en si mesma un sistema. O seu entorno é o sistema produtivo no que se insire do que recibe unha serie de entradas, en forma de demandas, e ó que entrega unha serie de saídas, ou oferta.

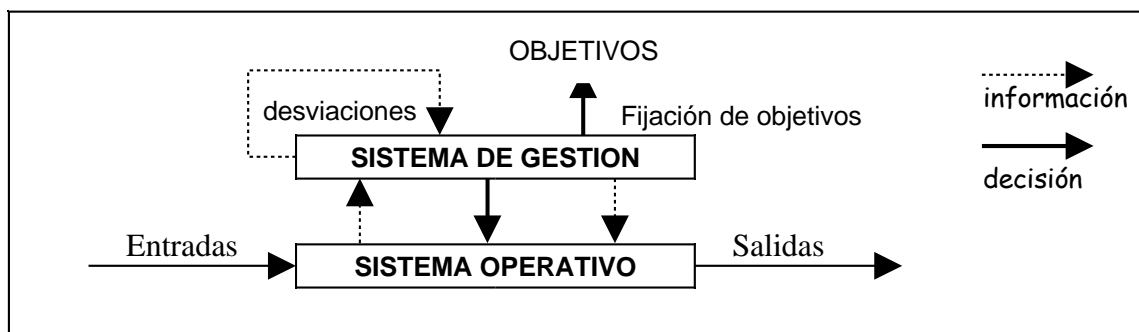
Á súa vez dentro da empresa ou organización existen varios subsistemas formados polos distintos departamentos ou áreas nas que se divide a empresa. Exemplo: Sistema produtivo da provincia de Lugo -> El Progreso -> Redacción-> Redacción de deportes.

Dentro das organizacións que serán obxecto do noso estudio podemos distinguir tres subsistemas:

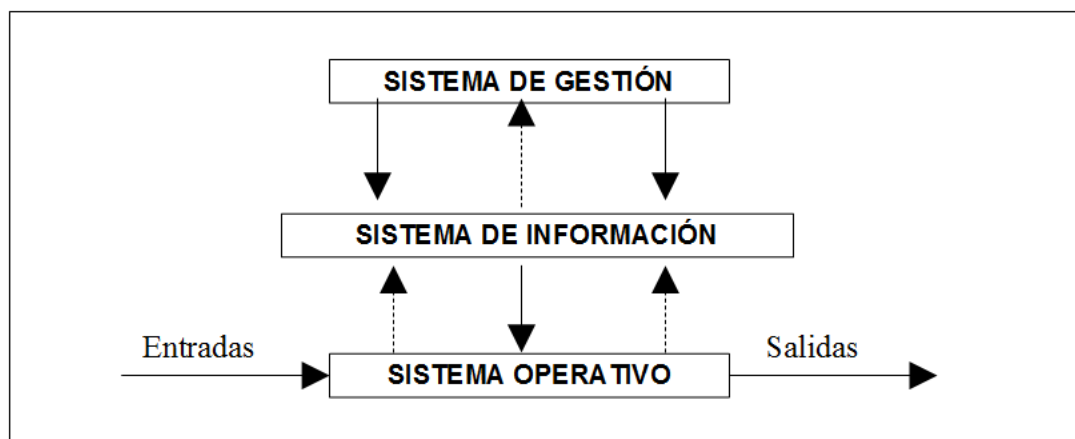
- O **sistema operativo** ou sistema físico. Transforma un fluxo de entradas (materias primas, fluxos financeiros, etc.) nun fluxo de saídas (produtos terminados, fluxos financeiros, ...).



- O **sistema de xestión** ou sistema de control. Responsable da fixación dos obxectivos e as estratexias a seguir para levalos a cabo, e que, ademais, se encarga de controlar o funcionamento do sistema físico e establecer as correccións necesarias nos obxectivos fixados en función da información que recibe deste. Estaría representado pola dirección da empresa.



- O **sistema de información**. É un sistema que actúa de intermediario entre o sistema operativo e o sistema de xestión.



Se tivéramos que resumir cunha soa frase o principal cometido dun Sistema de Información dentro dunha organización, poderíamos afirmar que este encárgase de entregar a información oportuna e precisa, coa presentación e o formato adecuados, á persoa que a necesita dentro da organización para tomar unha decisión ou realizar algunha operación e xusto no momento en que esta persoa necesita dispoñer de dita información.

Hoxe en día, a información debера ser considerada como un dos mais valiosos recursos dunha organización e o Sistema de Información é o encargado de almacenar e procesar a información que recibe dos outros subsistemas da organización e do exterior. A recollida e almacenamento da información realizarase de acordo cun modelo de datos que define que datos se deben almacenar e como se van a almacenar. Do bo funcionamento deste subsistema depende en gran parte o éxito global da organización.

Clasificación dos sistemas de información

O sistema de información existe en todas as empresas. Ata a aparición dos ordenadores o sistema de información funcionaba de xeito manual. Cando os ordenadores toman parte no funcionamento do sistema de información, falamos de sistema de información automatizado. No que queda deste apartado falaremos sempre destes últimos sistemas.

Polo xeral, **as clasificacións mais estendidas dos Sistemas de Información soen agrupar estes en función da súa finalidade.**

Dunha forma moi global pode considerarse que existen dúas funcións básicas para os sistemas:

- **Soporte ás actividades operativas**, que dá lugar a sistemas de información para axilizar actividades da empresa, tanto as máis estruturadas (aplicacións de contabilidade, nómina, pedidos e, en xeral, o que se denomina "xestión empresarial") como tamén sistemas que permiten o manexo de información menos estruturada: aplicacións ofimáticas, programas técnicos para funcións de enxeñería, etc.
- **Soporte ás decisións e o control de xestión**, que pode proporcionarse dende as propias aplicacións de xestión empresarial (mediante saídas de información existentes) ou a través de aplicacións específicas.

O soporte ás decisións require frecuentemente a consulta dos datos dunha maneira menos previsible que as aplicacións de xestión, por elo a tecnoloxía de Bases de Datos ten gran importancia para facer posible este tipo de aplicación. De feito, soe considerarse que a tecnoloxía de Bases de Datos permitiu a aparición do denominado Sistema de información para a xestión (MIS, polas súas siglas en inglés), que permiten abordar problemas non estruturados.

Os MIS utilizan os datos almacenados nos sistemas informáticos da empresa para xerar informes que permitan ós directivos mellorar o control de xestión das distintas áreas funcionais da empresa. Deste modo, conséguese axilizar o proceso de toma de decisións, ó proporcionar a información necesaria de forma rápida, precisa e fiable.

A dirección de alto nivel necesita ferramentas para diagnosticar problemas (análise) e para elixir a mellor alternativa (simulación, planificación ...). Este tipo de ferramentas requiren o manexo de información histórica e non estiveron dispoñibles ata principios dos anos noventa, dando lugar a un conxunto de aplicacións englobadas en termos como "aplicacións de soporte a decisións" (DSS), "software de apoio á dirección" (EIS, ESS), "Sistemas de Datawarehousing y Datamining" ou, de forma mais xenérica, Sistemas de "Intelixencia de Negocio" (Business Intelligence).

Sistemas de datawarehousing.

Os avances nas Tecnoloxías da Información fixeron posible o desenvolvemento dos Sistemas de Datawarehousing/Datamining, que constitúen o núcleo das aplicacións de Business Intelligence ('Intelixencia de Negocio').

Estes sistemas constan de tres elementos principais:

- Recollida e xestión de grandes volumes de datos: Tecnoloxía de Datawarehousing.
- Análise dos datos: Tecnoloxía OLAP e ferramentas de 'Minería de datos' (Datamining).
- Software de consulta amigable e intuitivo, alcanzable o usuario final.

Un Datawarehouse constitúe o elemento fundamental dun sistema de apoio á toma de decisións na empresa.

Podemos considerar que un Datawarehouse é un gran almacén de datos, no que se integran datos procedentes de varias fontes:

- Datos das distintas aplicacións de xestión da empresa (diseminados por distintos departamentos: administración, marketing, produción, etc.)
- Datos de fontes externas.

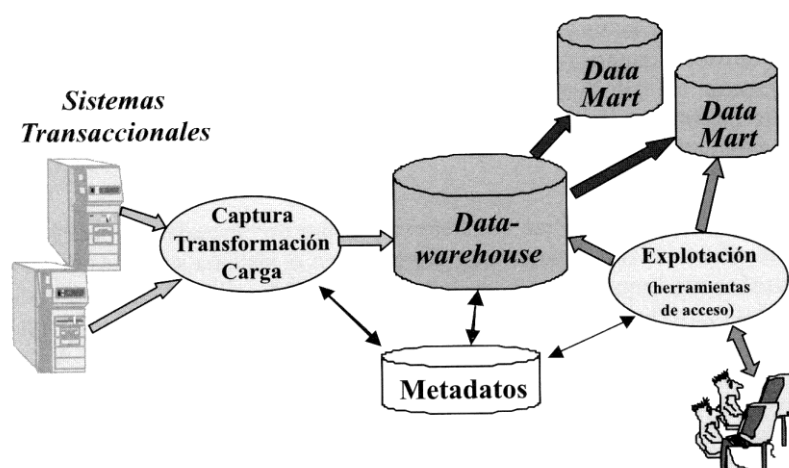
Entre as características dos datawarehouse atopámonos:

- Son de só lectura, os datos que se gardan neles non se borran nin modifican.
- O seu deseño está orientado a facilitar a lectura e o análise dos mesmos (OLAP, análise multidimensional). Isto fai que a súa estrutura non estea normalizada e admita redundancias.
- A estrutura dos almacéns de datos distingue entre “feitos” e “dimensións”. Os feitos virían a ser os datos en si que queremos analizar: as ventas, os pedidos dos clientes, etc. As dimensións serían os datos que organizan eses feitos segundo as características relevantes no análise: as provincias nas que se fan as ventas, a data das mesmas, o vendedor que as fixo.....
- Os datos gardados no datawarehouse reflicten a evolución temporal dos “feitos”.
- Algunhas das técnicas empregadas gardan datos agregados, isto é datos que corresponden a grupos de “feitos”. Os grupos fórmanse segundo os valores correspondentes das dimensións, polo que ó combinarse varias o número de grupos faise moi grande.
- Estas redundancias poden facer que o tamaño do datawarehouse medre exponencialmente, se se gardan datos agregados correspondentes a moitos grupos.
- Un aspecto importante na práctica son os sistemas e técnicas para capturar e gardar a información no almacén de datos.

Ademais, un Datawarehouse debe dispoñer dunha gran capacidade de almacenamento, xa que os datos pertencen a largos períodos de tempo (varios exercicios económicos da empresa).

Por outra parte, nestes sistemas, o diccionario de datos contén metadatos referidos á descrición e procedencia dos datos, as transformacións que experimentaron, a periodicidade da súa actualización, etc.

Nestes sistemas tamén se utiliza o concepto de Data Mart para referirse a un almacén de datos de menor tamaño, restrinxido xeralmente a un área temática ou a un departamento da empresa.



As ferramentas de datamining son técnicas avanzadas que permiten detectar e modelar relacións entre os datos e obter información non evidente: patróns de consumo, predición do comportamento dos clientes, asociacións de produtos, etc.

Cos Sistemas de Datawarehousing/Datamining, os directivos poden dispoñer da información necesaria en moi pouco tempo e co mínimo esforzo. Deste modo, os directivos poden dedicarse mais ó análise da información obtida e non tanto á súa busca.

Con estas ferramentas aplicadas ó marketing profúndase no coñecemento do comportamento dos clientes, o cal permite personalizar a oferta e conseguir unha maior eficacia das accións comerciais.

Exemplos de aplicación

Véxase algúns exemplos de aplicación dos Sistemas de Datawarehousing/Datamining levados a cabo por grandes empresas do noso país.

- No sector da banca, distintas institucións teñen utilizado estes sistemas dende finais dos anos noventa para conseguir unha mellor segmentación da súa carteira de clientes, a partir da análise dos produtos contratados e tipo de operacións realizadas. Isto permitiulles elaborar perfís de clientes para optimizar o envío de mailings con novas ofertas de produtos e campañas comerciais. E os resultados foron inmediatos: notable redución dos custes dos mailings e incremento das taxas de resposta.
- Por outra parte, a compañía Unión Fenosa empregou un Sistema de Datawarehousing e Datamining para construír un modelo de predición sobre potencial de compra da "Calor Económica" (sistema de calefacción que utiliza acumuladores de calor e a chamada Tarifa Nocturna").

En dito modelo analizáronse un conxunto de variables da súa base de datos de clientes: potencia contratada, consumo total anual, estacionalidade de consumos (relación entre o consumo de inverno e o consumo de verán), nivel de renda, tipo de municipio, dispoñibilidade de gas natural, etc.

Deste modo, a empresa conseguiu extraer 250.000 potenciais clientes da "Calor Económica" da súa base de datos global de mais de 2.400.000 clientes (datos referidos ó ano 1997). Convén destacar a redución de custes conseguida a partir desta análise de datos, xa que permitiu recortar á décima parte o esforzo comercial da campaña de "Calor Económico", sen que iso representara unha menor eficacia da campaña.

Na actualidade, no sector das telecomunicacións tamén teñen unha especial aplicación, sobre todo para predicir o comportamento de clientes e desenvolver novos tipos de servizos e de tarifas, adecuados ós patróns de comportamento destes clientes.

- Así, por exemplo, Telefónica Móviles puxo en marcha en 1998 o denominado "proxecto Minerva", un sistema de Datawarehousing para manexar grandes volumes de información procedentes das facturas e do rexistro de chamadas dos clientes (os destinos, a duración e os horarios), así como dos propios sistemas operacionais da compañía.

A partir da análise dos datos acumulados foi posible deseñar ofertas a medida e lanzar novos servizos (como o "friends and family", que ofrece tarifas especiais a uns números seleccionados polo cliente), determinar en que zonas xeográficas se producía un menor crecemento de aboados (para intensificar as campañas neses lugares), predicir as baixas de clientes, etc.

A operadora gastouse 4,21 millóns de euros neste sistema, integrado por servidores Alpha de 64 bits de Compaq cunha capacidade de almacenamento de 2 terabytes, polo software de bases de datos Oracle 8.04 Parallel Server e as ferramentas de análise e Datamining de ETI, Business Objects e Microstrategy. Segundo datos da propia empresa, reduciu durante o ano 1999 as baixas de clientes á metade grazas á utilización desta tecnoloxía de análise de datos.

- Outro exemplo destacado encontrámolo no xigante da distribución Wall Mart de Estados Unidos. Esta empresa detectou ás 9.30 horas do famoso 11 de setembro de 2001 que as vendas de bandeiriñas se incrementaran notablemente.

Grazas ó sistema informático que ten instalado puido reaccionar antes que a competencia e ás 10.30 horas pediu mais bandeiriñas ós subministradores, que llelas serviron ó día seguinte. A competencia tan só reaccionou un pouco mais tarde, á 12.30 horas, pero a capacidade de fabricación dos provedores fixo que estes lles puideran enviar a mercancía tres semanas despois, cando xa se pasara o momento.

1.1.2 Sistemas de ficheiros

Introdución

Podemos definir un ficheiro de datos como unha “estrutura de datos que se almacena nunha memoria secundaria do ordenador (disco, cinta, dvd, ...) consistente nun conxunto de informacións organizadas en unidades de acceso denominadas rexistros”.

Para comprender mellor o funcionamento dos ficheiros informáticos podemos utilizar un símil destes. Podemos comparalos cos típicos archivadores de fichas de cartón. A información de cada ficha equivale a un rexistro, e cada unha das informacións elementais que contén a ficha é o que chamaremos campo.

Por exemplo: Consideremos un archivador manual que contén as fichas cos datos dos empregados dunha empresa. Para cada empregado tense unha ficha que podería ter o formato seguinte:

Apelidos:	Nome:	NIF:
Domicilio:		
Poboación:	CP:	
Provincia:		
Data de nacemento:	Data de alta:	
Centro de traballo:		
Categoría:		
Posto de traballo:		
Estado civil:	Nº fillos:	Sexo(H/M):

O rexistro sería toda a información correspondente a un empregado. O campo sería cada unha das informacións elementais que se teñen de cada empregado, como por exemplo: apelidos, nome, categoría, ...

Para facilitar a busca dunha ficha determinada o mais normal é ter o ficheiro clasificado por algunha das informacións que aparecen na ficha e que teñen valores diferentes nos rexistros. Por exemplo, no archivador de empregados poderían estar ordenadas alfabeticamente polos apelidos, de forma que se queremos buscar a ficha dun empregado bastaría con saber os seus apelidos e buscariámolo na posición que lle corresponda por orde alfabética.

A estes campos que permiten identificar unha ficha e diferenciala das demais, facilitando o acceso a unha ficha determinada coñéceselles co nome de campo clave. A clave pode estar formada por mais dun campo, nese caso chámasele clave composta.

Nos ficheiros informáticos pódense utilizar mais dun campo clave, de forma que un deles considérase a clave principal ou clave primaria e ó resto das claves chámaseles claves secundarias ou claves alternativas. As claves alternativas son usadas habitualmente en procesos de busca. A clave primaria toma un valor diferente para cada rexistro, mentres que as

claves alternativas poden ter o mesmo valor en mais dun rexistro.

Como os ficheiros soen ser moi voluminosos só se poden levar á memoria principal partes deles para poder procesalos. A cantidade de información que é transferida entre o soporte no que se almacena o ficheiro e a memoria principal do ordenador nunha soa operación de lectura/gravación recibe o nome de rexistro físico ou bloque.

Normalmente en cada operación de lectura/gravación transfírense varios rexistros do ficheiro, é dicir un bloque soe conter varios rexistros. Ó número de rexistros que entran nun bloque coñéceselle co nome de factor de blocaxe, e a esta operación de agrupar varios rexistros nun bloque chámase bloque de rexistros.

É frecuente chamar volume ós soportes de almacenamento de datos que utiliza o ordenador, e tamén é normal que dentro dun volume se poidan almacenar varios ficheiros, polo que se fala de que este sería un volume multificheiro. En algunhas ocasións, os ficheiros que se utilizan en grandes empresas son tan grandes que non entran nun único soporte e faise necesario almacenalos en varios. A estes ficheiros chámase ficheiros multivolume.

Deseño de rexistros

Como vimos, un ficheiro é un conxunto de rexistros, e un rexistro está composto, á súa vez, por un conxunto de campos. Tendo en conta isto, no deseño dun ficheiro ten unha gran importancia o deseño dos rexistros que o forman, ademais da selección do soporte de datos mais adecuado para almacenalo e a forma en que se organizan os datos dentro do soporte.

Se observamos as fichas do arquivador de empregados, podemos ver que:

- A información referida a cada empregado está toda nunha ficha (rexistro).
- Sobre cada ficha as informacións (campos) aparecen nunha orde, que é a mesma para todas (estrutura do rexistro).
- As fichas están ordenadas por algún criterio, para facilitar a busca dunha ficha determinada. Por exemplo, polos apelidos, ou polo DNI (campo clave).

Tendo en conta todo isto, o deseño dun rexistro consiste en seleccionar, ordenar e definir as características dos campos que o forman. Normalmente para representar o deseño dos rexistros utilízanse unhas follas de deseño, nas que para cada campo se definen:

- Un **identificador** ou nome que nos permita identificalo nos programas.
- A **posición** do campo dentro do rexistro.
- O **tipo de datos** que vai conter (numérico, alfabético, lóxico,..)
- O **tamaño** do campo en caracteres (se o campo é numérico con decimais, debe indicarse o número de decimais que ten).
- Unha **descrición** do contido do campo, se fora necesario.

Unha vez que temos claro os campos que van a formar un rexistro, o maior problema que temos é seleccionar o tamaño de cada campo, debido a que a información que se vai gardar nun campo non é a mesma para todos os rexistros. Por exemplo, o campo APELIDOS do exemplo garda os apelidos dos empregados e o tamaño destes non son iguais para todos os empregados, podemos ter dende un DIAZ DIAZ, ata un RODRIGUEZ DE LA VILLAJOSYOSA GARCIA-DOMEQ.

Para solucionar este problema poden adoptarse varias solucións:

- Definir o campo co maior tamaño posible, o que pode supoñer un malgasto de soporte se un rexistro ten un valor moi longo con respecto ó resto.
- Definir o campo con un tamaño intermedio e abreviar os valores que superen ese tamaño. Esta solución é a mais aplicada, e utilízase sempre que se poida abreviar a información que almacena o campo.

- Utilizar campos con lonxitude variable para cada rexistro. Esta solución sería a que mellor aproveita o soporte, pero o seu manexo é complicado para o programador que tería que utilizar técnicas que lle permitan delimitar o comezo e o final de cada campo dun rexistro.
- Empregar un sistema de codificación da información utilizando unha táboa coa lista de posibles valores que pode tomar o campo e os códigos que lle corresponden. Por exemplo para o campo PROVINCIA, teríamos una táboa con todas as provincias e a cada unha delas asóciase un código de dous díxitos que se corresponda cos dous primeiros caracteres do código postal.

Selección do soporte

A estrutura de datos tipo ficheiro é unha estrutura externa, é dicir, que non se almacena na memoria principal do ordenador, elo é debido a que estas estruturas almacenan normalmente unha gran cantidade de información, que se vai a utilizar a longo prazo.

Os soportes mais utilizados para almacenar os ficheiros son os discos (magnéticos, ópticos, ou magneto-ópticos) e as cintas magnéticas. Dentro destes dous tipos de soporte existen no mercado unha gran variedade de modelos.

Todos estes soportes agrúpanse en dúas categorías en función do modo de acceso ós datos neles almacenados: mentres a cinta é un soporte secuencial, é dicir, se quero ler un dato que está na metade da cinta teño que ler todo o que hai ata chegar a esa posición; no disco o acceso ós datos pódese facer de forma directa podendo colocarnos na posición que nos interesa e ler a partir dela. Esta diferenza é moi similar á existente entre unha cinta de casete e un disco de música. Na cinta para escoitar a 3ª canción non queda mais remedio que facer pasar a cinta ata chegar a esa canción, pasando polas anteriores. Nun disco, se a canción que interesa é a 3ª so hai que pulsar o 3 e a cabeza de lectura desprazarase ó lugar en que empeza esta canción, sen necesidade de ter que pasar polas anteriores.

Normalmente a información que vai ser manexada con frecuencia almacenarase en discos, mentres que as cintas só se reservan case exclusivamente para as copias de seguridade.

Conceptos de organización e modo de acceso

O termo organización de ficheiros aplícase á forma en que se colocan os rexistros sobre o soporte informático (disco, cinta,...), no momento de seren gravados e como se localizan no momento de seren lidos.

Existen **dúas formas básicas de organización de ficheiros: secuencial e relativa.** Na **organización secuencial** os rexistros vanse gravando uns a continuación dos outros na orde que se van dando de alta, mentres que na **organización relativa** os rexistros grávanse nas posicións que se lles manda e que estarán en función do valor que garden no campo clave.

O modo de acceso refírese ó procedemento que se ten que seguir para poderse situar nun rexistro determinado co fin de facer unha operación de lectura ou gravación do mesmo.

O modo de acceso pode ser secuencial ou directo. No modo de acceso secuencial para chegar a un rexistro é necesario pasar por todos os anteriores, mentres que no modo de acceso directo pódese chegar directamente a un rexistro coñecendo unicamente o valor do campo clave.

Ó modo de acceso directo pódese chegar de varias formas:

- A posición que ocupa o rexistro dentro do ficheiro coincide co contido da clave.
- Calculando a posición que ocupa o rexistro no ficheiro mediante unha transformación do contido do campo clave (acceso aleatorio - Hashing).
- Mediante o uso de táboas de índices. A localización dun rexistro faise buscando na táboa de índices o valor do campo clave e obtemos a posición na que está gravado o rexistro dentro do ficheiro (acceso indexado).

A elección dunha forma de organización determinada está en función do tipo de aplicación que lle vamos a dar ó ficheiro. Dous factores inflúen directamente na selección: a memoria necesaria para almacenar os datos e a velocidade de acceso ós mesmos.

Exemplo: un ficheiro no que os rexistros se actualicen todos xuntos unha vez ó mes e o tempo necesario para a actualización non sexa importante, non pode ter a mesma organización que un ficheiro que se actualice todos os días varias veces poñendo ó día os datos que conteñen os seus rexistros e que ademais o tempo empregado para a actualización deba ser o menor posible.

Ficheiros con organización secuencial

Nun ficheiro con organización secuencial os rexistros vanse gravando sobre o soporte informático un a continuación de outro, sen deixar ocos no medio.

Neste tipo de ficheiros existe, polo tanto, unha correspondencia total entre a orde lóxica e a orde física. Orde lóxica é a orde na que os rexistros son dados de alta e recuperados, e orde física é a orde na que están gravados os rexistros no soporte.

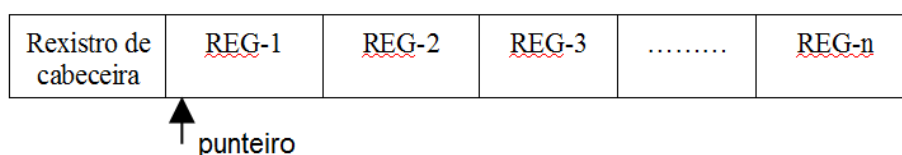
Os rexistros normalmente estarán ordenados polo contido dun ou mais campos para facer máis fácil o traballo de busca, inserción e borrado de rexistros.

A organización secuencial é aconsellable para ficheiros cun índice de utilización moi elevado e estables, é dicir, aqueles nos que se vaian a utilizar a maioría dos rexistros e que sufran poucas operacións de alta, baixa e modificación de datos.

Utilización de ficheiros con organización secuencial.

No momento de utilizar os ficheiros con organización secuencial temos que ter en conta o soporte sobre o que están gravados, pois algunhas operacións que se poden facer nos soportes direccionables non se poden facer nos soportes secuenciais.

A estrutura física dos ficheiros con organización secuencial podemos representala, dunha forma esquemática, así:



Un rexistro de cabeceira contén información para o sistema operativo acerca do ficheiro: posición na que está gravado o primeiro rexistro de datos, tamaño dos rexistros, posición en que está o final do ficheiro, etc...

Ó abrir o ficheiro, o sistema operativo utiliza un indicador ou punteiro para gardar o enderezo do primeiro rexistro de datos do ficheiro. Cada vez que se realiza unha operación de lectura ou gravación esa información actualízase co enderezo no que comeza o seguinte rexistro. Por esta razón, os rexistros só se poden gravar e ler secuencialmente.

Vantaxes e inconvenientes da organización secuencial

Podemos citar como vantaxes:

- Aproveita ó máximo o soporte, ó non deixar ocos entre os rexistros.
- Rápido acceso ó rexistro seguinte, polo que se fai ideal cando en cada operación de actualización ou consulta vanse a procesar a maioría dos rexistros.
- Pódense utilizar calquera tipo de rexistros: de lonxitude fixa, variable ou indefinida.
- Pódense gravar en calquera tipo de soporte, tanto secuenciais como direccionables.
- Todas as linguaxes de programación dispoñen de instrucións para traballar con este tipo de ficheiros.

Entre os inconvenientes destacaremos:

- O único modo de acceso é o acceso secuencial, polo que para ler o rexistro que ocupe a posición n é necesario ler os $n-1$ rexistros anteriores. Isto fai que este tipo de organización non sexa axeitada para ficheiros nos que se necesita procesar frecuentemente rexistros illados, é dicir que teñan un índice de utilización baixo.
- Non se poden inserir rexistros entre os que xa están gravados. Se temos o ficheiro ordenado polo contido dun campo e queremos dar de alta un rexistro, que segundo ese orden debería ir entre dous rexistros que xa existen no ficheiro, é necesario copiar todo o ficheiro nun novo, gravando no novo ficheiro todos os rexistros e inserindo o rexistro que se quere dar de alta na posición que lle corresponda. Outra posibilidade é ir dando de alta os rexistros ó final do ficheiro e a continuación realizar unha operación de reordenación de todo o ficheiro.
- Se o ficheiro está gravado nun soporte secuencial, por exemplo unha cinta magnética, para poder facer modificacións ou borrado de rexistros é necesario facer unha copia do ficheiro nun ficheiro novo. Se o soporte non é secuencial, as modificacións pódense facer sobre o mesmo rexistro, e o borrado pódese facer de forma lóxica, é dicir, mediante a gravación dunha marca no rexistro.

Variantes da organización secuencial.

Co fin de mellorar as prestacións da organización secuencial xorden unha serie de organizacións que son unha variante desta e que poden ser utilizadas con soportes direccionables. As mais empregadas son:

- A organización secuencial indexada, na que os rexistros cos datos grávanse nun ficheiro secuencialmente, pero pódense recuperar con acceso directo grazas á utilización dun ficheiro adicional, chamado de índices, que contén información da posición que ocupa cada rexistro no ficheiro de datos.
- A organización secuencial encadeada, que nos permite ter os rexistros ordenados segundo unha orde lóxica diferente da orde física na que están gravados, grazas á utilización duns campos adicionais que se introducen en cada rexistro chamados punteiros, que sinalan o enderezo do seguinte, ou anterior, rexistro en orde lóxica.

Ficheiros con organización secuencial-indexada.

Ós ficheiros con este tipo de organización chámaselles tamén ficheiros indexados porque baséanse na utilización de índices, que permiten o acceso a un rexistro do ficheiro de forma directa, sen ter que ler os anteriores. Estes índices son similares ós dos libros: se nos interesa ler un capítulo concreto podemos recorrer ó índice que nos di en que páxina comeza e

abrimos o libro por esa páxina, sen ter que mirar en todas as páxinas anteriores para localizalo.

As características mais relevantes dun ficheiro indexado son as seguintes:

- O deseño do rexistro ten que ter un campo, ou combinación de campos, que permita identificar cada rexistro de forma única, é dicir, que non poida haber dous rexistros que teñan a mesma información nel. A este campo chámasele campo clave e é o que vai a servir para crear o índice.

Un mesmo ficheiro pode ter mais dun índice asociado. De saber sabe que se vai a consultar con frecuencia o ficheiro polo contido dun campo, pode ser conveniente asociar un índice a ese campo.

- Permiten utilizar o modo de acceso secuencial e o modo de acceso directo para ler a información gardada nos seus rexistros.

O modo de acceso directo faise coñecendo o contido do campo clave do rexistro que queremos localizar. Con esa información o sistema operativo pode consultar o índice e coñecer a posición do rexistro dentro do ficheiro.

No modo de acceso secuencial os rexistros son lidos ordenados polo contido do campo clave, independentemente da orde na que se foron gravando (a orde lóxica non é igual á orde física) debido a que o acceso ós datos se fai a través do índice, que para facer mais fácil a busca dos rexistros permanece sempre ordenado polo campo clave.

- Soamente se pode gravar nun soporte direccionable. Como por exemplo o disco magnético. Se isto non fora así non podería empregar o acceso directo.

Estrutura dos ficheiros con organización secuencial-indexada

A estrutura dun ficheiro indexado varía de un fabricante a outro, pero, basicamente consta de:

- **Area de datos:** é a área na que se escriben os rexistros de datos cando se crea o ficheiro.
Os rexistros dun ficheiro con organización secuencial-indexada grávanse nun soporte de almacenamento directo, en secuencia ascendente, de acordo cos valores da clave e en páxinas ou bloques de lonxitude fixa.
- **Area de índices:** é creada polo sistema ó mesmo tempo que se van almacenando os datos. Contén unha táboa que asocia as claves cos enderezos dos rexistros na área de datos.
Cada entrada da área de índices está formada polo valor mais alto da clave de cada grupo de rexistros e un punteiro co enderezo do primeiro rexistro do grupo.
- **Area de desbordamiento:** tamén chamada de excedentes, ou de overflow, onde se gravan os rexistros que non teñen sitio na área de datos. Os novos rexistros insírense e quedan enlazados entre si mediante punteiros, conservando a orde lóxica en función do campo clave ou índice principal.

Do tratamento dos índices e punteiros encárgase o sistema operativo polo que non vai a crear problemas ó programador cando manexa este tipo de ficheiros. O usuario sabe o que sucede cando solicita unha consulta dun rexistro, pero non sabe como se realiza internamente esa consulta.

Índices a varios niveis.

Cando o ficheiro de datos ten moitos rexistros, pode que o ficheiro de índices se faga demasiado longo e as buscas nel se vaian facendo lentas, xa que cantos mais rexistros teña, mais medra o tempo necesario para localizar un rexistro. Para evitar este problema recórrase á utilización de índices a varios niveis, de forma que se substitúe unha área de índices única,

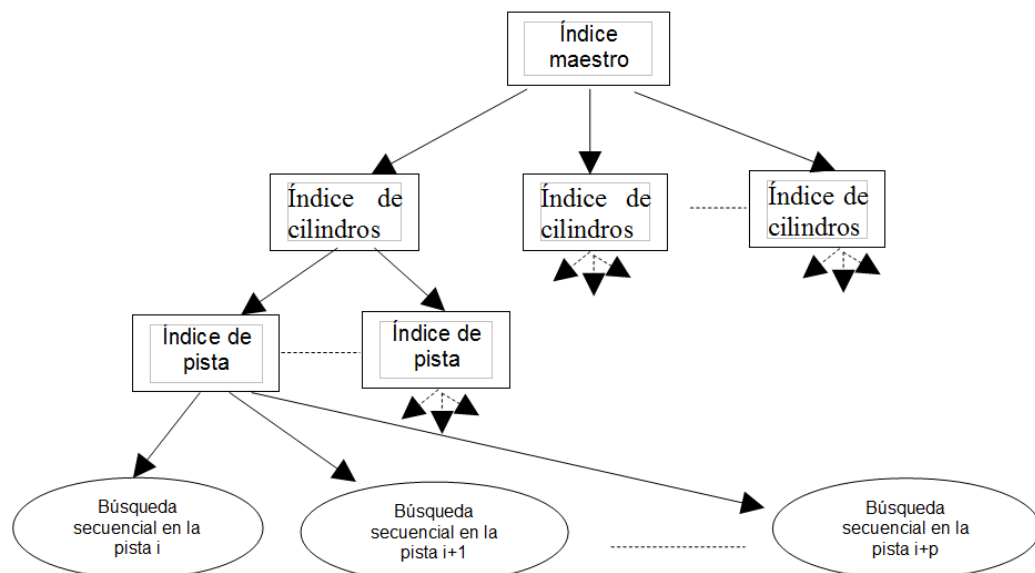
pero dun tamaño grande, por varias áreas de índices de menor tamaño onde a busca é mais rápida.

Existen diferentes formas de manexar os índices dependendo do sistema operativo ou a linguaxe de programación coa que se traballa. As organizacións de índices mais utilizadas na práctica son a ISAM (Indexed Sequential Access Method) e a VSAM (Virtual Storage Access Method). A implementación varía dun fabricante a outro dependendo da estrutura de datos empregada para almacenar o índice e dos algoritmos empregados na manipulación de esa estrutura. A maioría das implementacións ISAM empregan como estrutura de almacenamento unha estrutura de árbore, nas súas diversas variantes, con un sistema de índices de múltiples niveis.

Para darnos unha idea de como se organizarían os índices a varios niveis podemos empregar un exemplo dun ficheiro, con organización ISAM, gravado nun disco duro cun certo número de pratos. Podemos representar de forma moi simplificada a posible estrutura do almacenamento do arquivo, tendo en conta que unha “pila de discos” está composta de cilindros, pistas e sectores. A estrutura de índices ISAM podería ter estes niveis:

- 1º Nivel: (Índice mestre). Sinala o cilindro no que se encontra o rexistro que nos interesa.
- 2º Nivel: (Índice de cilindros) Sinala en que pista está ese rexistro, dentro do cilindro.
- 3º Nivel: (Índice de pista) Sinala, dentro da pista, en que bloque ou sector se encontra o rexistro.

No seguinte gráfico vemos como podería ser o algoritmo de busca dun rexistro nun ficheiro con unha organización ISAM con este tipo de estrutura de datos:



Vantaxes e inconvenientes da organización secuencial indexada.

A organización secuencial indexada permite o acceso directo ós rexistros sen os inconvenientes da existencia de ocos como na organización directa e o tratamento secuencial dos rexistros propio da organización secuencial.

Esta organización é moi utilizada tanto para procesos nos que interveñen poucos rexistros como para aqueles nos que se manexa o ficheiro completo.

Vantaxes:

- Permite o acceso secuencial. No acceso secuencial, ademais, os rexistros lense ordenados polo campo clave.

- Permite o acceso directo ós rexistros. Realmente emula o acceso directo, empregando para elo as táboas de índices. Primeiro busca a clave na área de índices e logo vai ler á área de datos no enderezo que lle indica a táboa.
- Permite ter o ficheiro ordenado por varios campos, e polo tanto, o acceso directo pode ser tamén por distintos campos. Cada campo clave leva asociada un área de índices.

Inconvenientes:

- Ocupa mais espazo no disco que os ficheiros secuenciais, debido ó uso da área de índices.
- Ten tendencia a que aumente o tempo medio de acceso ós rexistros cando se producen moitas altas novas con claves que hai que intercalar entre as existentes, xa que aumenta a área de overflow. Require unha reorganización periódica.
- So se pode utilizar en soportes direccionables.

1.2 Bases de datos

1.2.1 Introducción

Os primeiros sistemas de procesamento de datos estaban pensados para executar as tarefas administrativas; para “reducir o papeleo”. Mais recentemente, os sistemas evolucionaron cara a produción e xestión da información, que se converteu nun recurso vital para as empresas. Se os directivos dispoñen dunha boa información, é probable que poidan tomar decisións acertadas e con unha gran rapidez.

O desenvolvemento dos sistemas de bases de datos converteuse en crucial para proporcionar información correcta (sen erros) e oportuna (no momento en que se necesita) ós directivos como axuda á toma de decisións.

Os sistemas informáticos tradicionais foron chamados por algúns autores sistemas orientados cara o proceso debido a que neles ponse o énfase nos tratamentos que reciben os datos, os cales almacénanse en ficheiros deseñados para unha determinada aplicación. As aplicacións analízanse e implántanse con enteira independencia unhas de outras, e os datos non se soen transferir entre elas, se non que se duplican sempre que os correspondentes traballos os necesitan.

Debido a esta independencia no deseño das aplicacións, pode ocorrer que un mesmo dato se recolla en varios ficheiros, manexados por distintas aplicacións. Esta redundancia produce, ademais dunha ocupación inútil de memoria secundaria, un aumento dos tempos de proceso ó repetirse os mesmos controles e operacións nos distintos ficheiros.

Pero mais graves aínda son as inconsistencias que a miúdo se presentan nestes sistemas, debido a que a actualización dun dato, cando se encontra en máis dun ficheiro, non se soe realizar de forma simultánea en todos os ficheiros que o conteñen.

Os problemas son aínda mais acusados cando se presentan demandas inesperadas de información. Se non estaban previstas nos programas das aplicacións habería que deseñar un novo programa que permitise facer esa consulta sobre os ficheiros de datos e incorporalos ás aplicacións que os necesitan.

Desta análise dedúcese claramente a necesidade dunha xestión mais racional do conxunto de datos, xurdindo así un novo enfoque que se apoia na utilización de bases de datos, nas cales os datos son recollidos e almacenados unha soa vez, con independencia do número de aplicacións que os manexan.

Vemos, polo tanto, que a solución dos problemas asociados ó tratamento dos datos nos sistemas tradicionais leva a un cambio radical no enfoque do sistema de información, no cal “os datos organízanse e mantéñense nun conxunto estruturado que non está deseñado para

unha aplicación concreta, se non que, pola contra, tende a satisfacer as necesidades de información de toda a organización”; necesidades cuxa diversidade vese acentuada co transcurso do tempo.

Estes sistemas orientados cara os datos van substituíndo, de forma progresiva, ós sistemas orientados cara o proceso que pola súa pouca fiabilidade, falta de adecuación á realidade, baixa aportación ós novos sistemas de xestión, e mal asegurada confidencialidade foron perdendo de forma progresiva a confianza dos usuarios.

1.2.2 Concepto de base de datos

Unha base de datos é un conxunto, colección ou depósito de datos estruturados, que están almacenados nun soporte informático de acceso directo.

Dada a importancia que teñen no mundo real as interrelacións entre os datos, é imprescindible que a base de datos sexa capaz de almacenar estas interrelacións ó igual que fai con outros elementos, sendo esta unha diferenza esencial respecto ós ficheiros onde non se almacenan as interrelacións.

A redundancia dos datos debe ser controlada de forma que non existan duplicidades prexudiciais nin innecesarias, e que as redundancias físicas, convenientes moitas veces a fin de responder a obxectivos de eficiencia, sexan tratadas polo propio sistema de xestión, de modo que non poidan producirse incoherencias. Polo tanto, un dato actualízase lóxicamente polo usuario de forma única, e o sistema preocuparase de cambiar fisicamente todos aqueles campos nos que o dato estivese repetido, en caso de existir redundancia física.

A actualización e recuperación nas bases de datos debe realizarse mediante procesos ben determinados, incluídos nun conxunto de programas que se encargan da xestión da base de datos e que se denominan sistemas xestores de bases de datos (S.G.B.D); procesos que deben de estar deseñados de modo que se manteña a integridade, seguridade e confidencialidade da base de datos.

O concepto de base de datos foi cambiando e configurándose ó longo do tempo. Na actualidade, e de acordo con estas características que acabamos de analizar, podemos definir a base de datos como:

“ Colección ou depósito de datos integrados, con redundancia controlada e cunha estrutura que reflecta as interrelacións e restricións existentes no mundo real. O datos, que han de ser compartidos por diferentes usuarios e aplicacións, deben manterse independentes de estas, e a súa definición e descrición, únicas para cada tipo de datos, han de estar almacenadas xunto cos mesmos. Os procedementos de actualización e recuperación, comúns e ben determinados, haberán de ser capaces de conservar a integridade, seguridade e confidencialidade do conxunto dos datos”.

1.2.3 Vantaxes das bases de datos fronte ós ficheiros clásicos

As bases de datos, xurdidas como resposta ó novo plantexamento dos sistemas orientados cara os datos, para mellorar a calidade das prestacións dos sistemas informáticos e aumentar o seu rendemento, presentan unha multitude de **vantaxes fronte ós sistemas clásicos de ficheiros, debido, sobre todo, a que se basean nunha estrutura de datos integrada e centralizada, eliminando así os problemas de redundancia e control dos datos. As vantaxes dos sistemas de bases de datos son, entre outras, as seguintes:**

- **Independencia dos datos respecto ós tratamentos e viceversa.** Os programas non conteñen a descrición da estrutura de almacenamento dos datos, polo que a inclusión de novos datos na base de datos, desaparición de outros, cambios na estrutura física, etc., non deben obrigir a modificar os programas, evitando o importante esforzo que orixina a reprogramación das aplicacións.

- **Coherencia dos resultados.** Ó non existir (ou a lo menos diminuír en gran medida) a redundancia nos datos, desaparece o problema que se presentaba no enfoque clásico, de que o cambio dun dato obrigaba a actualizar unha serie de ficheiros. Desta forma elimínase tamén o inconveniente das diverxencias nos resultados debidas a actualizacións non simultáneas en todos os ficheiros.
- **Maior eficiencia na recollida, validación e entrada dos datos ó sistema.** Ó non existir apenas redundancias, os datos recóllense e válídanse unha soa vez, aumentando así o rendemento de todo o proceso previo ó almacenamento. Ademais, estas operacións non se deixan unicamente en mans das aplicacións, se non que en última instancia é o Sistema Xestor de Bases de Datos (SXBD) o que se encarga da validación.
- **Mellor dispoñibilidade dos datos para o conxunto dos usuarios.** Cando se aplica a tecnoloxía de bases de datos, cada usuario ou aplicación xa non é propietario dos datos posto que estes compártense entre o conxunto de aplicacións, existindo unha mellor dispoñibilidade dos datos para todos os que teñen necesidade e estean autorizados para o seu acceso. Ademais, os usuarios poden facer consultas non previstas polas aplicacións utilizando o linguaxe de consultas ou outras ferramentas propias do SXBD
- **Redución do espazo de almacenamento.** A desaparición (ou diminución) das redundancias, así como a aplicación de técnicas de compactación, leva ós sistemas de bases de datos a unha menor ocupación de espazo no soporte.

1.2.4 Compoñentes dun sistema de base de datos

Un sistema de base de datos é algo mais que simples datos ou que un conxunto de datos en combinación cuns programas de xestión. Un sistema de base de datos está formado polos seguintes compoñentes:

- **Datos.** A principal característica destes sistemas é que a información debe estar integrada e compartida:
 - Integrada: a Base de Datos pode considerarse como un conxunto de ficheiros de datos que son tratados como un so e no que se eliminou totalmente, ou en parte, a redundancia de datos.
 - Compartida: Os datos poden compartirse entre varios usuarios distintos. É posible que varios destes usuarios accedan ó mesmo tempo ó mesmo elemento de información (acceso concorrente).
- **Equipo (Hardware).** Conxunto de dispositivos físicos utilizados para almacenar e procesar os datos.
 - Ordenadores, utilizados para procesar os datos da Base de datos: poden ser mainframe, miniordenador ou ordenador persoal. O mainframe e os miniordenadores foron utilizados tradicionalmente para soportar o acceso de varios usuarios a unha base de datos común. Os ordenadores persoais eran empregados, inicialmente, para bases de datos autónomas manexadas por un usuario único. Non obstante, actualmente solen estar conectados a unha rede baseada nunha plataforma cliente/servidor, garantindo o acceso de varios usuarios a unha base de datos común almacenada en unidades de disco dun servidor, e controladas por este. O servidor pode ser outro ordenador persoal mais potente, ou ben, un miniordenador ou un mainframe.
 - Volumes de almacenamento. Xeralmente son unidades de disco que constitúen o mecanismo de almacenamento principal para as bases de datos.
 - Outros dispositivos, como unidades de cinta, terminais, impresoras, etc.
- **Programas (Software).** Un sistema de base de datos inclúe dous tipos de programas:
 - O software de propósito xeral, para a xestión da base de datos, chamado normalmente Sistema Xestor de Bases de Datos (SXBD, ou tamén DBMS, en inglés). O SXBD

manexa todas as solicitudes de acceso á base de datos formuladas polos usuarios e os programas de aplicación.

- O software de aplicación, que usa as facilidades do SXBD para manipular a base de datos co fin de levar a cabo unha función específica na xestión da empresa (por exemplo: a xestión de almacén). Pode estar desenvolvido nunha linguaxe de programación estándar, tal como COBOL, Visual Basic, Java ou C, ou ben, nunha linguaxe propia dos SXBD denominadas linguaxes de cuarta xeración (4GL).
- **Persoal.** Nun sistema de base de datos interveñen un número importante de usuarios, que podemos clasificar en tres grupos:
 - Administrador da base de datos (ABD). Son os encargados de deseñar a estrutura da base de datos e os responsables de que o sistema funcione correctamente. O ABD encárgase de deseñar a estrutura da base de datos, autorizar o acceso á base de datos, de coordinar e vixiar a súa utilización, etc. O ABD é o responsable cando xorden problemas como violacións de seguridade ou unha resposta lenta do sistema.
 - Programadores de aplicacións, que se encargan de desenvolver as aplicacións que manexan datos da base de datos. Estas aplicacións conterán solicitudes de datos ó SXBD que logo serán procesados polos programas da aplicación que terán como finalidade resolver problemas específicos da empresa.
 - Usuarios finais, que son persoas que non teñen por que ter coñecementos informáticos e que poden manipular os datos (examinalos e actualizalos) coa axuda das aplicacións, ou ben de linguaxes de consulta non procedementais (non é necesario indicar o algoritmo de acceso ós datos), tipo SQL, ou ben, mediante ferramentas que forman parte do SXBD

1.2.5 Modelos de datos

Modelar consiste en definir un mundo abstracto e teórico tal que as conclusións que se poidan sacar del coinciden coas manifestacións aparentes do mundo real.

Chamamos modelo ao instrumento que se aplica a unha parcela do mundo real (universo do discurso UD) para obter unha estrutura de datos á que denominamos esquema. O modelo será un conxunto de conceptos, regras e convencións que nos permiten describir os datos do UD.

Os modelos serven para describir a estrutura dunha base de datos, é dicir os tipos de datos, as relacións entre eles e as restricións que deben cumprir.

Os modelos son a base para as linguaxes de datos. Por exemplo, o SQL é o resultado de aplicar unha determinada sintaxe ao modelo relacional.

Aos valores que toman os distintos obxectos dun esquema nun momento determinado denominarémolos instancia ou ocorrencia do esquema.

Os modelos que contaron ata hai pouco con maior grao de aceptación son tres:

- Modelo xerárquico.
- Modelo en rede.
- Modelo relacional.

Na actualidade o modelo relacional é o máis usado pero a complexidade crecente das aplicacións informáticas actuais está a facer que o modelo de bases de datos orientadas a obxectos gañe cada vez máis adeptos e, pola contra, os modelos xerárquico e en rede están a deixarse de utilizar.

Modelo xerárquico

Un modelo de tipo xerárquico utiliza árbores para a representación lóxica dos datos e ten as seguintes características:

- Os rexistros (chamados segmentos) están dispostos en forma de árbore e non poden existir ciclos.
- Os rexistros só poden estar enlazados mediante relacións uno a un ou un a moitos.
- Cando se elimina un rexistro pai débense borrar todos os seus rexistros fillos.

Modelo en rede

Os modelos en rede baséanse na utilización dunha estrutura non lineal en rede (ou Plex), na que cada rexistro fillo pode ter máis dun nodo pai.

O modelo en rede máis estendido é o modelo CODASYL que presenta a restrición de que non permite directamente a relación moitos a moitos. Se se presenta, utilízase un rexistro intermediario, chamado conector ou liga-D, que contén os campos clave principais dos rexistros que se desexan relacionar.

Modelo relacional

Este modelo é posterior aos dous modelos anteriores, e foi desenvolvido por Codd en 1970.

Un modelo relacional utiliza táboas bidimensionais (relacións) para a representación lóxica dos datos e as relacións entre eles.

Chámaselle rexistro, entidade ou tupla a cada fila da táboa e campo ou atributo a cada columna da táboa. Unha clave será un atributo ou conxunto de atributos que identifique de forma única a unha tupla.

As táboas deben cumprir unha serie de requisitos:

- A táboa só pode ter un tipo de rexistro.
- Non existen rexistros duplicados.
- Os rexistros dentro dunha relación non teñen unha secuencia determinada.
- Pódense crear novas táboas relacionando campos procedentes de dúas ou máis táboas xa existentes.

Modelo orientados a obxectos

O modelo orientado a obxectos baséase en encapsular código e datos nunha única entidade chamada obxecto. O interface entre un obxecto e o resto do sistema defínese mediante un conxunto de mensaxes.

Un obxecto ten asociado:

- Un conxunto de variables que conteñen os datos do obxecto.
- Un conxunto de mensaxes aos que o obxecto responde.
- Un método que é un anaco de código para implementar cada mensaxe.

1.2.6 Tipos de bases de datos.

En función do modelo de datos empregado

- Bases de datos xerárquicas. Seguen o modelo de datos xerárquico. Un exemplo de bases de datos xerárquica é IMS de IBM desenvolvido a finais dos 60.
- Bases de datos en rede. Seguen o modelo de datos en rede. Exemplos de bases de datos en rede son : ADABAS, TOTAL, IMAGE.
- Bases de datos relacionais. Seguen o modelo de datos relacional. Actualmente os sistemas relacionais son un estándar no mercado, especialmente en operacións comerciais. Exemplos de sistemas de bases de datos relacionales son: Informix, DB2, Oracle, SYBASE, Ingress, SQLServer, MySQL, PostgreSQL...

En función da localización da información

- Bases de datos centralizadas. Os datos almacénanse nun ordenador. Poden atender a varios usuarios pero o software de xestión da base de datos (SXBD) e a base de datos almacénanse nunha única máquina.
- Bases de datos distribuídas. A base de datos está almacenada en varios ordenadores conectados en rede. Poden ser:
 - Replicadas: mantéñense copias de toda a base de datos, ou partes dela, en dous ou mais servidores, permitindo que certos datos da base de datos estean almacenados en mais dun sitio e así aumentar a dispoñibilidade dos datos e mellorar o rendemento das consultas globais.
 - Fragmentadas: o sistema divide a base de datos en varios fragmentos e garda cada fragmento nunha máquina diferente.

1.3 Outros sistemas de almacenamento

Ademais das bases de datos existen outros sistemas de almacenamento da información como poden ser os servizos de directorios e XML.

1.3.1 Servizos de directorio

Un Servizo de Directorio (SD) é unha aplicación ou un conxunto de aplicacións que almacena e organiza a información sobre os usuarios e recursos dunha rede de ordenadores, e permite ós administradores xestionar o acceso de usuarios ós recursos sobre dita rede.

Pode considerarse como unha base de datos especializada na que o número de lecturas é moi superior ó de escrituras. Por isto, están optimizados para almacenar información relativamente estática, polo que non son recomendables para almacenar datos que cambian con frecuencia.

As características fundamentais dos servizos de directorios son:

- Intentan optimizar as buscas e lecturas, mentres que non importa que por elo se penalicen as actualizacións.
- As transaccións e as operacións de restauración xeralmente non están implementadas.
- Os datos solen ser redundantes, sendo o seu principal obxectivo as buscas eficientes.
- Os datos organízanse nunha estrutura xerárquica.

1.3.2 XML

XML é o acrónimo de *eXtensible Markup Language*. XML é un formato estándar do *World Wide Web Consortium* (W3C) deseñado a partir de SGML (Standard Generalized Markup Language, lenguaje de marcado generalizado estandar) para representar datos estruturados de forma xerárquica. Os documentos XML inclúen unha serie de etiquetas que permiten crear documentos autocontidos, nos que os datos van sempre acompañados dos seus metadatos correspondentes. É unha linguaxe de marcas extensible (metalinguaxe) pensada para a transmisión de información estruturada e considérase un estándar para o intercambio de datos entre aplicacións independente do formato de almacenamento dos mesmos.

XML é unha tecnoloxía sinxela que ten o seu redor outras que a complementan e fana moito máis grande e cunhas posibilidades moito maiores. Ten un papel moi importante na actualidade xa que permite a compatibilidade entre sistemas para compartir a información dun xeito seguro, fiable e fácil.

XML está provocando a aparición de novas tecnoloxías, entre elas, a aparición dunha nova xeración de bases de datos, as chamadas bases de datos nativa XML ou *native XML database* que aínda que se atopan nunha fase de investigación e desenvolvemento, nun futuro poden ser unha boa alternativa ás xa coñecidas bases de datos relacionais. Este tipo de bases de datos son completamente distintas ás relacionais, as cales na actualidade teñen soporte para XML, pero aínda seguen almacenando toda a información de xeito relacional, é dicir en forma tabular (táboas, rexistros e columnas), pero a principal característica que brindan estas bases de datos é a capacidade de obter os resultados das consultas en formato XML; é por iso que ditas bases de datos pertencen á categoría de *XML-enabled database*.