

Administració de les dades

Jordi Serra Ruiz
Miquel Colobran Huguet
Josep Maria Arqués Soldevila
Eduard Marco Galindo

PID_00190186



Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>

Índex

Introducció	5
Objectius	6
1. Les dades i l'organització	7
2. On és la informació	10
3. La consulta de la informació	13
3.1. Les consultes de la direcció	14
3.2. Servidors de bases de dades	15
3.3. ERP	16
3.4. Magatzem de dades	16
3.4.1. Arquitectura d'un magatzem de dades	18
4. Protecció de la informació	21
4.1. Seguretat de la xarxa	21
4.2. Còpies de seguretat	21
4.3. Seguretat en bases de dades	22
4.3.1. Confidencialitat de la informació	23
4.3.2. Disponibilitat de la informació	23
4.3.3. Integritat de la informació	24
5. Tasques/responsabilitats de l'administrador	25
Resum	26
Activitats	27
Exercicis d'autoavaluació	27
Solucionari	28
Glossari	29
Bibliografia	30

Introducció

Actualment un dels grans valors de totes les organitzacions és la informació (també en podríem dir dades, tot i que ja veurem quina diferència hi ha). Aquesta és la matèria primera del sistema informàtic. Per tant, és el que s'ha de conèixer més bé on és, perquè com que tots els ordinadors estan connectats per les xarxes informàtiques, correm el perill que les dades estiguin disperses per tota l'organització en les estacions de treball. Si això passa, no sabrem ni quines dades tenim ni on les hem d'anar a buscar.

Necessitarem saber on és la informació per a fer-ne còpies de seguretat, ja que en cas que hi hagi un desastre el programari es pot reinstal·lar, però les dades les ha creat l'organització, no es poden "comprar" enlloc. En ser un dels actius més importants actualment, cal protegir-la.

També és important saber on són aquestes dades perquè les podem combinar per a obtenir dades noves sobre la nostra organització.

Per a la direcció (i en general per a tota l'organització), disposar de la informació adient a temps permet de prendre decisions correctes en cada situació en el moment oportú.

Objectius

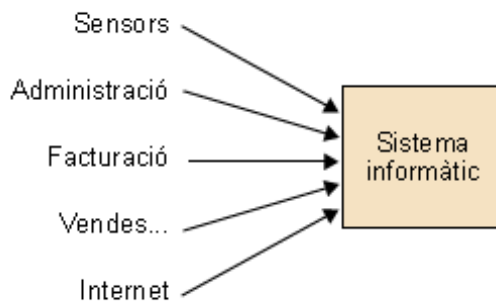
Els materials d'aquest mòdul contenen les eines necessàries perquè l'estudiant assoleixi els objectius següents:

- 1.** Saber que la informació de l'organització és molt valuosa.
- 2.** Comprendre que és necessari saber on hi ha tota la informació de l'organització.
- 3.** Saber diferents maneres de mantenir la integritat de la informació.
- 4.** Saber per què s'han fet populars els servidors de bases de dades i altres arquitectures de dades.
- 5.** Saber en quins llocs hi pot haver les dades de l'organització i quins són els millors llocs on podrien ser.
- 6.** Conèixer els fonaments de seguretat de les bases de dades.

1. Les dades i l'organització

Una organització crea dades contínuament, per la qual cosa una manera de considerar el sistema informàtic és com si només es tractés d'un magatzem de dades.

Sistema informàtic com a magatzem de dades



Aquest “gran dipòsit” de dades en brut no és útil d'aquesta manera, perquè si les dades no tenen una organització i una coherència, amb la gran quantitat que n'hi ha, no tindria cap mena de sentit.

Abans de continuar endavant, hem de dir que hi ha una diferència entre informació i dades. La definició formal de cadascun d'aquests conceptes és la següent.

Les **dades** són els registres dels successos. La **informació** és el processament de les dades perquè tinguin sentit.

Exemple sobre la diferència entre dada i informació

70293 pot voler dir 7 de febrer de 1993 o embalatge 7029 per al camió número 3. Un altre exemple: 1813 pot ser un número de matrícula de cotxe o pot ser una hora, les 18 hores i 13 minuts.

Informació i dades

Malgrat la diferència formal entre dades i informació, sovint els informàtics utilitzem aquestes dues paraules com a sinònimes. Com que el processament per a donar sentit a les dades es pot fer en molts llocs, nosaltres també farem servir indiferentment els termes *dades* i *informació*.

Generalment es guarden dades i es presenta informació a l'usuari, és a dir, que el processament d'aquesta informació perquè tingui sentit es fa en el programari en el moment de presentar-la en el dispositiu de sortida. En principi això és el que dona més flexibilitat al sistema, ja que permet de convertir les dades a qualsevol format de sortida. Tanmateix, sovint això queda lluny de la realitat, perquè diferents programaris guarden les dades amb diferents formats. A més a més, la majoria de programaris són incompatibles entre ells, per la qual cosa la recuperació creuada d'informació no és gens senzilla.

Així, doncs, les organitzacions creen constantment grans quantitats de dades i el que fan els sistemes informàtics és processar-les i distribuir-les entre tots els elements de l'organització per a augmentar l'eficàcia del conjunt.

El sistema informàtic pretén guardar la informació de l'organització perquè sigui fàcil recuperar-la posteriorment i de la mateixa manera que s'havia guardat.

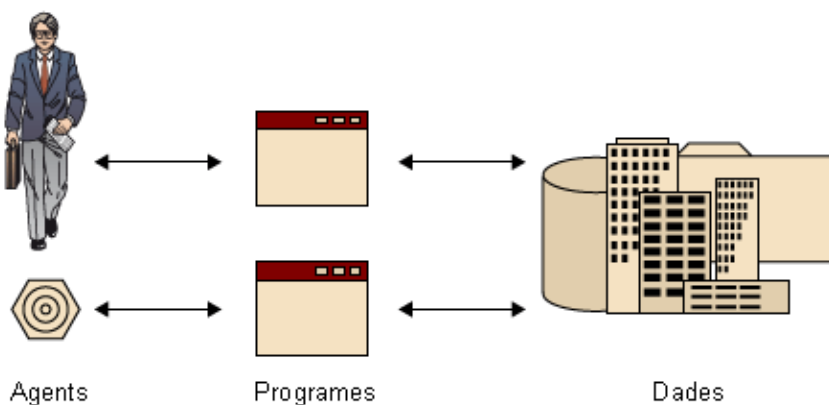
El primer objectiu del sistema informàtic és guardar la informació però fent-ne possible la recuperació igual que s'havia guardat. Ara bé, amb això desaprofitem el gran potencial dels ordinadors: la capacitat de manipular informació.

En aquest context, "manipular informació" té un sentit força ampli. Aquests són alguns dels significats possibles:

- Poder afegir informació nova a la que ja teníem.
- Poder esborrar informació obsoleta de la que ja teníem.
- Poder modificar informació incorrecta de la que ja teníem.
- Poder relacionar diferents informacions per a construir informació nova, contrastar la coherència o la veracitat de la informació que es vol entrar, reduir la redundància de la informació o buscar-hi noves relacions.

La interacció de les dades amb l'organització la podem veure d'aquesta manera:

Esquema d'interacció de dades amb l'organització



Els **agents** poden ser, per exemple, persones o sensors que a partir de programes accedeixen a les dades de l'organització. També podríem pensar en aquests agents com a entitats que estan fora de l'organització, a Internet, però que interaccionen amb el sistema informàtic i modifiquen d'alguna manera les dades del sistema.

Els **programes** s'utilitzen per a accedir a les dades de l'organització d'una manera ordenada i correcta. Sembla lògic que l'accés a les dades, atès que són un bé valuós de l'organització, s'ha de protegir. Els programes s'encarreguen (juntament amb el sistema operatiu) de fer tots els controls necessaris per a evitar els accessos no desitjats a les dades de l'organització.

Amb aquesta visió de les dades de l'organització, sembla senzill, viable i potser fins i tot fàcil resoldre el problema d'un usuari que ens vingui a dir que necessita extreure determinada informació del sistema informàtic. Pel que hem vist aquí, fins ara, aquest problema es reduiria a crear un nou programa que accedís a les dades. De moment, per exemple, no hem tingut en compte com es manegen els permisos i els grups que incorporen els diferents fitxers que integren la informació de l'organització per a protegir el contingut, ja que la informació s'ha de protegir d'accessos indiscriminats.

El sistema informàtic recull informació, la guarda, la manipula i la presenta a l'usuari.

2. On és la informació

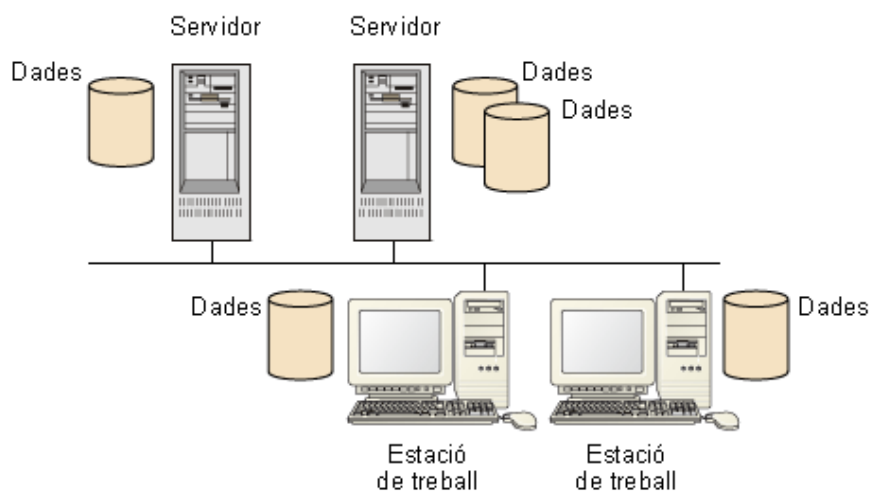
Desgraciadament, la realitat no és tan simple. En l'apartat anterior tot era ideal: guardàvem les dades i, quan les necessitàvem, les agafàvem i les presentàvem (imprimíem). Si era el que ens calia, ja havíem acabat. La realitat involucra molts més processos que fan que el sistema sigui més complex. Aquests són alguns dels problemes:

- Les dades estan guardades en fitxers.
- Les dades estan guardades en un format determinat.
- Les dades no sempre estan guardades on es necessiten.
- Les fonts de dades, és a dir, els sensors, les persones, els formularis web, etc. que generen dades, les generen en un format determinat.
- Cada font de dades pot fer servir un format diferent, i es pot guardar en fitxers potencialment diferents i en ubicacions diferents.

El que ens passa en la pràctica és que les dades de l'organització estaran disperses des del punt de vista del programari. No tots els programes podran accedir a totes les dades, independentment dels permisos de l'usuari.

El problema és encara més greu si ens fixem en l'estructura de la xarxa. Vegem els llocs on hi pot haver potencialment informació. Les dades poden estar en aquests llocs perquè s'ha decidit així, o perquè els usuaris, per desconeixement, generen informació i la guarden en aquests llocs, sense consultar-ho.

Esquema d'una xarxa



Informació relacionada?

Formularis web diferents poden generar fitxers de text diferents, i un sol usuari, segons l'aplicació que utilitzi, crearà informació diferent, en formats diferents, que anirà a parar a fitxers diferents. Si, per exemple, l'usuari fa un document amb processadors de textos del balanç anual comptable per al banc, i ho fa en diferents fitxers segons els comptes, crearà informació relacionada (del mateix tema) que és molt difícil de poder relacionar mai l'una amb l'altra.

Podem tenir informació en qualsevol directori de qualsevol partició de qualsevol disc dels servidors o de les estacions de treball. Fins i tot en ordinadors portàtils.

És a dir, anant a l'extrem, podem tenir informació duplicada en màquines diferents, amb sistemes operatius diferents, que tinguin sistemes de fitxers diferents.

El pitjor és que part d'aquesta informació pot ser crítica i molta pot ser desconeguda pel departament d'informàtica, o potser sap que existeix i no la té controlada i, per tant, no disposa de cap mecanisme de recuperació davant d'un desastre.

Tenim, doncs, un problema que pot arribar a ser molt greu i que s'ha repetit històricament en totes les organitzacions. Vegem possibles solucions:

- A partir de la taula d'aplicacions i del disseny dels usuaris en general, sabem quin programari utilitzen els usuaris. També sabem, a partir de la taula d'aplicacions, on hi ha la informació. El manteniment de la que estigui en els servidors és responsabilitat de l'administrador d'usuaris o de l'administrador de servidors. En qualsevol cas, sabem on és i disposem dels mecanismes de seguretat, protecció i recuperació adients en cas d'emergència.
- A partir de la taula d'aplicacions i del disseny dels usuaris en general, sabem quin és el programari que genera informació en local. De tot aquest programari, s'ha d'implementar el mecanisme de còpies de seguretat que faci falta, i habilitar els permisos i la seguretat que siguin necessaris per a garantir-ne la confidencialitat i poder-les incorporar al mecanisme general de còpia per a estar previngut en cas de fallada.
- Dins de la categoria d'informació "no controlable" hi acostuma a haver el programari d'ofimàtica (fulls de càlcul, processadors de textos, petites bases de dades, agendes, etc.). S'ha de posar especial atenció en el fet que les dades siguin a la unitat de xarxa privada de l'usuari (de la qual es fa una còpia de seguretat cada dia).
- Igualment, s'ha de formar els usuaris en l'ús de les eines informàtiques amb què treballen perquè aprenguin a posar la informació en la seva unitat personal de xarxa en lloc de fer-ho en la unitat local (o gravar-ho on proposa l'aplicació per defecte, ja que aleshores moltes vegades ni el mateix usuari no sap on hi ha realment guardades les dades). Aquesta "cultura informàtica" pot evitar la dispersió de la informació per tots els discos de l'organització, que és un dels grans problemes d'administració.

Vegeu també

Recordeu que hem tractat com es fa el manteniment de la informació dels usuaris en el mòdul "Administració d'usuaris".

Informàtica portàtil

Actualment, amb el creixement en les organitzacions de la informàtica portàtil, hi ha un gran problema afegit al de la dispersió, ja que aquests equips necessiten autonomia i funcionen sense estar connectats a la xarxa de l'organització. La primera qüestió que se'ns planteja és: què passa amb les dades que necessita per a treballar? Poden ser sensibles si surten fora de l'organització sense cap tipus de protecció. La segona qüestió és que aquests equips modifiquen o afegixen dades al sistema informàtic (estan desconectats de la xarxa). Com sincronitzen aquesta informació amb el sistema de l'organització? Una darrera qüestió seria que de vegades necessiten funcionar connectats a la xarxa de l'organització com si fossin una estació de treball més, i de vegades els cal connectar-se a la xarxa de l'organització des de fora. Com es pot fer això?

S'han de buscar maneres de saber on hi ha tota la informació de l'organització. Ens servirà, per exemple, per a garantir-ne la seguretat.

3. La consulta de la informació

De la mateixa manera que l'organització genera una gran quantitat de dades constantment, els usuaris també necessiten nova informació de l'organització molt sovint. Això pot voler dir noves consultes a les dades o generar noves fonts de dades en el sistema. Per tant, l'administrador ha de conèixer totes les bases de dades de l'organització.

Ja que la seguretat de tota la informació de l'organització és responsabilitat de l'administrador, ha de:

- Conèixer-ne la localització.
- Disposar d'un mètode de restauració en cas de problemes. Això implica còpies de seguretat, plans de continuïtat de negoci, etc.
- Tenir una idea general del contingut. Ha de tenir informació de la informació (en podríem dir metainformació).

La localització és important perquè té una gran rellevància en les polítiques de còpies de seguretat i en els temps d'accés per als usuaris, com també la facilitat d'accés, d'ampliacions futures, d'ampliació del grup d'usuaris que accedeixen a les dades, etc.

La idea general del contingut té utilitat per a l'optimització del sistema (no s'han de repetir bases de dades ni dades que ja en formen part) i per a les peticions de la direcció, que generalment van encaminades a extreure informació del sistema.

Quan aconseguim tenir totes les dades en els servidors (algunes particions, d'alguns discos, d'alguns servidors) el problema no està resolt, però almenys hem avançat molt. Ja podem fer algunes coses:

- Que diferents grups/usuaris accedeixin a les mateixes dades (les compareixin). Això els permet d'afegir-hi elements i consultar, modificar i esborrar aquestes dades.
- Com que les dades són als servidors, la seguretat és molt més alta.
 - Còpies de la informació
 - Pèrdua per fallada de l'equip
 - Robatori
 - Permisos i, per tant, restriccions d'accés

El terme *base de dades*

En aquest cas fem servir el terme *base de dades* en el seu sentit més ampli, ja que engloba des d'un full de càlcul fins a una base de dades completa, passant fins i tot per un document d'un processador de textos. Tot és informació de l'organització.

Mineria de dades

Els mètodes de mineria de dades pretenen extreure informació coherent a partir de quantitats d'informació no coherent i probablement dispersa. Hi ha una assignatura que tracta d'aquest "problema".

- Més facilitat d'accés i de difusió, perquè en aquestes condicions la informació és potencialment accessible per a tota l'organització (està controlat per la seguretat).

3.1. Les consultes de la direcció

També sembla que ens podem començar a plantejar un problema del qual fins ara no havíem parlat: el problema de la direcció.

La direcció veu el sistema informàtic com un magatzem d'informació i fa consultes a aquesta informació per a tenir una ajuda en la presa de decisions.

Per tant, per a prendre decisions fan peticions del tipus “quin és el producte que ha donat menys hores de producció i més benefici de vendes?”, o “quin producte ha costat menys de vendre als comercials?”, o “quins productes tenen menys cost de transport?”, o “de quins components hem tingut més pèrdues?”, o bé, finalment, per exemple, “quins treballadors han rendit menys?”.

Totes són preguntes sovint força difícils de contestar, malgrat que l'administrador pot arribar a saber que la informació és dins el sistema.

Abans, amb les dades distribuïdes per tota l'organització, era impossible contestar cap d'aquestes consultes. Ara, com que les dades són als servidors, són totes accessibles. Quin és el problema?

Per a contestar aquesta pregunta cal tornar una mica enrere i veure com s'arriba al punt actual sobre la informació de l'organització. El procés de guardar/recuperar informació el podem esquematitzar de la manera següent:

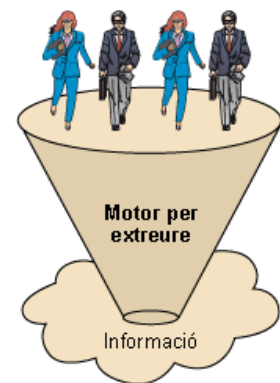
Esquema del flux d'informació en una organització



Input: Sensors, persones, aplicacions, etc.

Output: Llistats.

De moment només hem aconseguit solucionar el pas B, ja que tenim tota la informació accessible dins els servidors. Col·lateralment hem millorat el problema de seguretat i de compartició de la informació, però el que no hem millorat gens és el punt A: la informació continua entrant amb formats completament heterogenis, és en una gran quantitat de fitxers diferents i, per tant,



La direcció necessita extreure informació

els programes per a manipular aquesta informació poden ser extremadament complexos de fer (per no dir impossibles). Si manipular informació en formats heterogenis per a obtenir informació coherent és complex, i la direcció demana consultes sovint, el problema esdevé intractable.

Si, finalment, obtenim algun resultat d'alguns d'aquests fitxers amb formats heterogenis, aquest resultat també estarà en un format heterogeni (la mateixa informació guardada de dues maneres). Què s'ha de fer per a fabricar la llista final (pas D)?

No n'hi ha prou de tenir tota la informació en els servidors per a poder dir que ja la podem manipular com vulguem.

3.2. Servidors de bases de dades

Davant d'aquest conflicte apareix en el mercat un primer intent per a solucionar-lo, els servidors de bases de dades. La idea és molt simple. Integrar tota la informació en un sol lloc. Aquest lloc ha de ser ràpid, segur, fàcil d'utilitzar, fàcil de fer créixer, fàcil de fer-hi còpies de seguretat i ha d'incorporar permisos.

Els servidors de bases són sistemes gestors de bases de dades (SGBD¹), tot i que actualment també es coneixen com a *servidors per a bases de dades*, perquè funcionen amb tecnologia client/servidor, és a dir, que les estacions de treball obtenen la informació de les bases de dades fent peticions d'informació a un servidor a través de la xarxa.

⁽¹⁾SGBD és la sigla catalana que es correspon amb l'expressió anglesa *relational database management system* (RDBMS).

Sistemes gestors de bases de dades

En el mercat hi ha molts sistemes gestors de bases de dades (SGBD) comercials, i alguns de gratuïts. En general, tots demanen grans quantitats de disc i de memòria RAM. Si les bases de dades han de ser grans, llavors cal una unitat de control de procés (CPU) ràpida per a poder processar les peticions SQL i les transaccions ràpidament. Si es preveu que el servidor anirà carregat amb força informació, és molt important seguir les indicacions del fabricant de la base de dades sobre els requisits de l'ordinador en relació amb el disc, la RAM i el CPU per a obtenir-ne un bon rendiment.

Aquests programaris incorporen interfícies per a poder interrogar o programar les bases de dades des de qualsevol llenguatge, i una vegada s'ha creat la base de dades (base de dades, taules, estructura, camps, etc.) es pot fer servir des de qualsevol aplicació que hagi estat preparada per a fer-ho. Fins i tot des d'un servidor web amb extensions.

Aquest conjunt de característiques ha fet que moltes organitzacions hagin migrat les seves aplicacions cap a d'altres d'orientades a servidors de bases de dades. D'aquesta manera obtenim els avantatges següents:

- Tota la informació està concentrada en els servidors.

- Tota la informació, malgrat que pugui entrar en formats heterogenis, està guardada en format homogeni.
- És possible relacionar informació de fonts diferents, perquè està guardada en un sol lloc i en un mateix format.
- Permet de fer sortides molt més fàcilment.
- Permet de fer petites modificacions d'informes molt més fàcilment.
- Possibilita tenir tot el conjunt de dades de l'organització molt més integrat.
- Simplifica molt les còpies de seguretat.
- Facilita les consultes de la direcció.

Hi ha una solució per a tenir tota la informació en un sol lloc, un servidor de bases de dades.

3.3. ERP

Actualment encara s'ha dut molt més enllà la idea dels servidors de bases de dades. Sobre un servidor de bases de dades s'instal·len integrades les aplicacions que sabem que utilitzen totes les organitzacions. Després es comercialitza el producte.

El resultat s'anomena sistema de planificació de recursos de l'empresa (ERP²). És un conjunt de mòduls o paquets (les aplicacions). Per exemple, comptabilitat, facturació i nòmina tot perfectament integrat i funcionant sobre un servidor de bases de dades.

Aquests sistemes de gestió de la informació, en integrar i automatitzar molts aspectes operacionals de l'organització, homogeneïtza tot el procés d'introduir informació i els informes. En integrar tota la informació en un mateix lloc les consultes que es poden crear són moltes i molt potents.

Les grans empreses, amb milers de punts de treball i diferents seus, fan servir majoritàriament aquest tipus d'eines ja que garanteixen l'accés a les dades de manera més eficient.

3.4. Magatzem de dades

Amb els SGBD³ i els ERP⁴ solucionem el pas B i una part del pas C del flux d'informació en una organització, però no hem solucionat el problema de les consultes de la direcció. En especial perquè sovint les consultes fan referència

Ofimàtica i informàtica mòbil

L'ofimàtica i informàtica mòbil hi continuen essent, per la qual cosa no tota la informació és dins el servidor de bases de dades. Per tant, s'ha d'educar els usuaris en l'ús de les unitats personals situades en els servidors com a mesura per a protegir i poder compartir, si cal, la informació.

⁽²⁾ERP és la sigla d'*enterprise resource planning*.

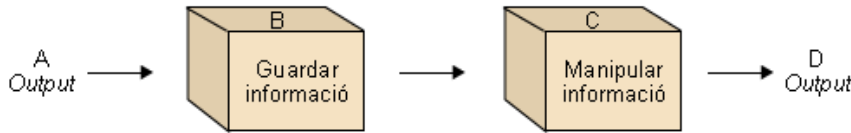
⁽³⁾SGBD és la sigla de *sistemes gestors de bases de dades*.

a intervals de temps que les nostres bases de dades no preveuen. Per exemple, les preferències dels nostres clients durant els darrers tres estius cap a quins productes s'han decantat?

⁽⁴⁾ERP és la sigla d'*enterprise resource planning*, en català 'sistema de planificació de recursos de l'empresa'.

Recordem l'esquema següent:

Esquema del flux d'informació en una organització



Input: Sensors, persones, aplicacions, etc.
Output: Llistats.

Els magatzems de dades⁵ són un pas més en l'àrea de l'emmagatzemament de dades. En lloc d'estar orientats a la manipulació de dades (altes, modificacions, baixes...), conegudes com a dades operacionals, estan orientats a guardar les dades per a fer cerca i consultes.

⁽⁵⁾Magatzem de dades en anglès s'expressa com a *data warehouse*.

Un magatzem de dades és una col·lecció de dades orientada a un determinat àmbit, integrat, no volàtil i variable en el temps, que ajuda la presa de decisions en l'organització en què s'utilitza.

Podem analitzar la definició de magatzem de dades:

- **Orientat a temes.** Les dades a la base de dades estan organitzades de manera que tots els elements de dades relatius al mateix esdeveniment o objecte del món real quedin units entre si.
- **Variable en el temps.** Els canvis produïts en les dades al llarg del temps queden registrats perquè els informes que es puguin generar reflecteixin aquestes variacions.
- **No volàtil.** La informació no es modifica ni s'elimina. Quan s'ha emmagatzemat una dada, aquesta es converteix en informació de només lectura, i es manté per a futures consultes.
- **Integrat.** La base de dades conté les dades de tots els sistemes operacionals de l'organització, i aquestes han de ser consistents.

Un magatzem de dades, per tant, és una base de dades corporativa que es caracteritza per integrar i depurar informació d'una o més fonts diferents, per a després processar-la permetent-ne l'anàlisi des de moltes perspectives. Pot estar dissenyat sobre un SGBD, però no són el mateix.

SGBD	Magatzem de dades
Dades operacionals	Dades de l'organització per obtenir informació
Orientat a l'aplicació	Orientat al tema
Actual	Actual i històric
Detallada	Detallada i resumida
Canvia constantment	Estable

Alguns dels avantatges d'implantar un magatzem de dades són:

- Proporciona una eina per a la presa de decisions en qualsevol àrea funcional, basant-se en informació integrada i global de l'organització.
- Facilita l'aplicació de tècniques estadístiques d'anàlisi i modelització per trobar relacions ocultes entre les dades del magatzem.
- Permet d'aprendre de les dades del passat i de predir situacions futures en diversos escenaris.
- Simplifica dins de l'organització la implantació de sistemes de gestió integral.

3.4.1. Arquitectura d'un magatzem de dades

L'arquitectura bàsica d'un magatzem de dades i els seus elements és la següent:

- **Bases de dades operacionals.** Bases de dades que registren les transaccions necessàries per a un correcte funcionament de l'organització. Aquestes dades són la font principal per al magatzem de dades.
- **ETL.** Periòdicament, s'han d'importar dades al magatzem de dades. És necessari normalitzar les dades abans d'introduir-les al magatzem de dades mitjançant eines d'extracció, transformació i càrrega (ETL⁽⁶⁾). Aquestes eines llegeixen les dades primàries (bases de dades OLTP⁽⁷⁾ de l'organització), realitzen el procés de transformació al magatzem de dades (filtratge, adaptació, canvis de format...) i escriuen al magatzem.

⁽⁶⁾ETL són les sigles en anglès d'extreure, transformar i carregar' (*extract, Transform and Load*).

⁽⁷⁾OLTP és la sigla en anglès de 'processament de transaccions en línia' (*online transaction processing*). Són els SGBD.

- **Magatzem de dades.** El dipòsit on s'emmagatzemen totes les dades de l'organització.
- **Metadata.** Un dels components més importants de l'arquitectura d'un magatzem de dades són les metadades. Es tracta de dades que descriuen quina és l'estructura de les dades que s'emmagatzemaran i com es relacionen. Les metadades documenten, entre altres coses, quines taules hi ha en una base de dades, quines columnes té cada taula i quin tipus de dades es poden emmagatzemar. Les dades són d'interès per a l'usuari final i la metadada és d'interès per als programes que han de manejar aquestes dades.
- **Data Mart.** Un subgrup lògic del magatzem de dades complet. Els Data Mart són subconjunts de dades per a ajudar que una àrea específica de l'organització pugui prendre millors decisions. Les dades existents en aquest context poden ser agrupades i explorades de moltes maneres per diversos grups d'usuaris segons les seves necessitats. Es podria dir que els Data Mart són petits magatzems de dades centrats en un tema o una àrea específica dins de l'organització.
- **Aplicacions d'usuari.** Són un conjunt d'eines que fan les consultes, analitzen i presenten la informació. Sovint es parla d'un client de magatzem de dades. Es poden trobar diversos tipus d'aplicacions d'usuari.
 - Eines de consulta i creació d'informes.
 - Eines de desenvolupament d'aplicacions.
 - Sistemes d'informació executiva (EIS⁸).
 - Eines de mineria de dades.
 - Sistemes de suport a la presa de decisions (DSS⁹).

Observació

Sobre els Data Mart es poden construir EIS (*executive information systems*; en català, 'sistemes d'informació per a directius') i DSS (*decision support systems*; en català, 'sistemes d'ajuda a la presa de decisions').

⁽⁸⁾EIS són les sigles de l'expressió anglesa *executive information system*.

⁽⁹⁾DSS són les sigles de l'expressió anglesa *decision support system*.

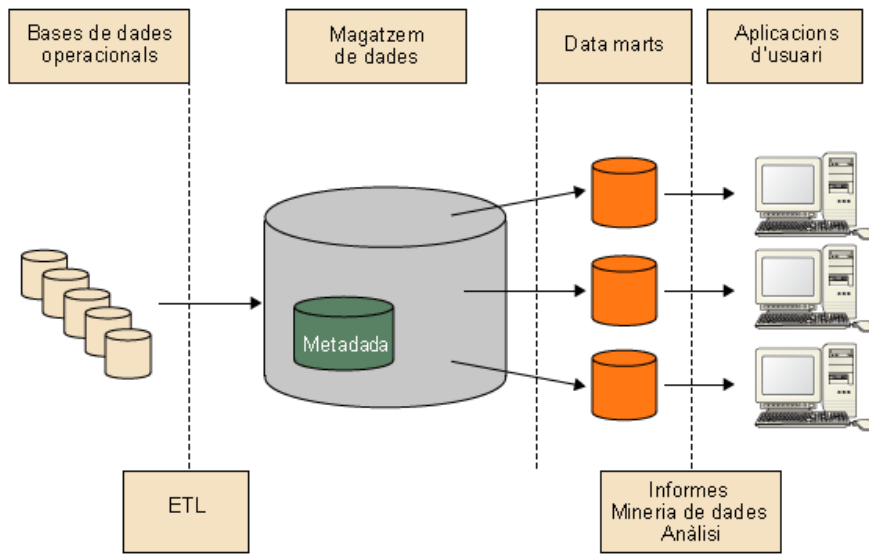
Modelització multidimensional

Per a la consulta i el modelat de dades en el magatzem de dades, s'utilitza el modelat multidimensional, que és una alternativa als models entitat-relació.

Mineria de dades

Es coneix com a *minería de dades* (en anglès, *data mining*) el procés d'anàlisi de grans quantitats de dades amb l'objectiu d'extreure informació útil. Per exemple, per realitzar classificacions o prediccions.

Esquema de l'arquitectura bàsica d'un magatzem de dades



4. Protecció de la informació

La protecció de la informació es pot mirar des de dos vessants. En el primer, senzillament es tracta de protegir-la d'accessos no desitjats. En el segon, es tracta de protegir-la per a no perdre-la.

La informació és un dels béns més valuosos de l'organització. S'ha d'invertir una part important dels esforços per a protegir-la. Tan problemàtic és perdre totes les dades en un incendi com que uns atacants accedeixin a les dades privades dels empleats de l'organització. A més, la llei de protecció de dades obliga a mantenir aquestes informacions en sistemes segurs.

4.1. Seguretat de la xarxa

La informació és dins la xarxa de l'organització, en servidors, en estacions de treball, en portàtils, etc. El personal hi accedeix, i és amb els mecanismes d'accessos que es limita i es controla qui pot accedir a aquesta informació, de quina manera i què hi pot fer. Un disseny de l'entorn d'usuaris correcte és el primer pas per a evitar accessos no desitjats a la informació.

L'altre pas per a tenir sistemes segurs és evitar intrusions. És molt difícil tenir sistemes completament segurs; per tant, només podem mirar de fer el sistema tan segur com sigui possible.

4.2. Còpies de seguretat

Les còpies de seguretat de les dades d'una organització és un dels temes més importants dins de l'administració de les dades, ja que sense una bona política de recuperació davant de desastres la informació es pot perdre. Bàsicament, s'ha de tenir present que les dades s'haurien de copiar a diari d'una manera íntegra, i tenir un historial de còpies de seguretat, que segons l'organització pot anar des de pocs mesos fins a anys.

Si s'ha seguit una política que reforça que la informació sigui als servidors les còpies són molt més senzilles, ja que poden estar perfectament integrades dins el procés normal de còpia dels servidors.

La política per a reforçar que es guardi la informació en el servidor passa per fer accions com, per exemple, les següents:

Vegeu també

Vegeu com es fa el disseny de l'entorn dels usuaris en el mòdul "Administració d'usuaris".

Vegeu també

Vegeu maneres d'optimitzar la seguretat del sistema en el mòdul "Administració de la seguretat".

Vegeu també

Vegeu les polítiques de còpies de seguretat en l'apartat 6 del mòdul "Administració de servidors".

- Que les aplicacions guardin per defecte en els espais personals de l'usuari del servidor. D'aquesta manera, si l'usuari no diu explícitament on vol guardar la informació, anirà al servidor.
- Formar l'usuari en l'entorn informàtic, explicar-li l'avantatge de guardar la informació en la unitat de xarxa com, per exemple, el canvi d'ordinador per trasllat de lloc de treball o la pèrdua de dades en cas de fallada de l'ordinador.
- Formar l'usuari en les aplicacions que ha de fer servir per a ensenyar-li que ho ha de guardar tot en unitats de xarxa.
- Fer que totes les aplicacions corporatives guardin la informació en les unitats de xarxa.

4.3. Seguretat en bases de dades

Podem veure una base de dades com un “conjunt de dades integrades, adequat a diversos usuaris i a diferents usos”. De manera que els problemes de seguretat vindran per l'ús simultani d'aquestes dades.

La protecció de les dades s'ha de fer contra fallades físiques, fallades lògiques i errades humanes (siguin o no intencionades), que poden alterar o corrompre les dades i ocasionar que la base de dades esdevingui inútil per al que es va crear.

Qualsevol SGBD ha de proporcionar tècniques que permetin que els usuaris tinguin accés únicament a una part de la base de dades i no a la resta. De manera que els SGBD tenen un subsistema de seguretat d'autorització encarregat de garantir la seguretat de parts de la base de dades contra l'accés no autoritzat.

Les bases de dades porten mecanismes per a prevenir fallades (subsistema de control), per a detectar-los quan s'ha produït (subsistema de detecció) i per a corregir-los després que s'ha detectat (subsistema de recuperació).

Els aspectes fonamentals de la seguretat en les bases de dades són:

- Confidencialitat. No s'han de proporcionar dades a usuaris no autoritzats. Inclou aspectes de privadesa (protecció de dades personals).
- Accessibilitat. La informació ha d'estar disponible.
- Integritat. S'ha d'assegurar que les dades no han estat falsejades.

4.3.1. Confidencialitat de la informació

Per a facilitar l'administració, els SGBD incorporen el concepte de perfil, rol o grup d'usuaris que reuneix un conjunt de privilegis. De manera que l'usuari assignat a un grup hereta tots els privilegis del grup.

El subsistema de control d'accés s'encarrega de denegar o concedir l'accés als usuaris. En un SGBD poden haver-hi diversos tipus d'autorització:

- **Autorització explícita.** Usada en els sistemes tradicionals. Es tracta d'emmagatzemar qui pot accedir a quins objectes de la base dades amb quins privilegis. S'acostuma a utilitzar una matriu de control d'accessos.
- **Autorització implícita.** L'autorització sobre un objecte es pot deduir a partir d'altres. Per exemple, si es pot accedir a una classe en un SGBD, es pot accedir a totes les instàncies de la classe.
- **Autorització forta.** Quan les autoritzacions deduïdes no poden ser invalidades.
- **Autorització dèbil.** Es permeten excepcions sobre les autoritzacions implícites.
- **Autorització positiva.** Si hi és, indica l'existència de l'autorització.
- **Autorització negativa.** És la negació explícita d'una autorització.

Recordeu que prèviament ja hem definit la política de seguretat de l'entitat: quins empleats poden accedir a les diferents fonts d'informació que hi ha a l'empresa i quins no. Això es definirà conjuntament entre el cap d'informàtica i la persona responsable de les dades o usuaris.

4.3.2. Disponibilitat de la informació

Els sistemes de bases de dades han d'assegurar la disponibilitat de les dades a aquells usuaris que hi necessiten l'accés. Així doncs, hi ha mecanismes que permeten recuperar la base dades contra fallades lògiques o físiques que destrueixen totalment o parcialment les dades.

El principi bàsic que sustenta la recuperació de la base de dades davant de qualsevol fallada és la redundància física. Les fallades més habituals són les provocades per fallades elèctriques, fallades del maquinari i fallades en els dispositius d'emmagatzematge (discos).

Utilitats de seguretat

Els SGBD contenen utilitats pròpies, però des d'un punt de vista de seguretat, s'hauria de pensar en alternatives com, per exemple, servidors tolerants a fallades.

Aquesta situació va motivar l'aparició del concepte de **transacció**. Davant de qualsevol fallada, s'ha de poder assegurar que, després d'una actualització, la base de dades quedi en un estat consistent. Per a aconseguir-ho es creen unes unitats d'execució anomenades transaccions, que es poden definir com a seqüències d'operacions que s'han d'executar de forma atòmica. O es realitzen totes les operacions de la transacció globalment o no se'n fa cap. Així es pot recuperar la base de dades de manera segura, ja que les operacions es fan per grups, que podem tornar a llençar en cas de fallada d'algun element de la base de dades.

4.3.3. Integritat de la informació

En aquest context, s'entén per integritat, la correcció, validesa o precisió de les dades de la base de dades. L'objectiu és protegir la base de dades contra operacions que puguin introduir inconsistències a les dades. El subsistema d'integritat d'un SGBD ha de detectar i corregir tant com sigui possible les operacions incorrectes.

Hi ha dos tipus d'operacions que poden violar la integritat de les dades: les operacions semànticament inconsistents i les interferències degudes a accessos concurrents.

- **Operacions contra la integritat semàntica.** Hi ha operacions que poden vulnerar restriccions definides en el disseny de la base de dades (per exemple, restriccions sobre els dominis o sobre els atributs). Aquestes restriccions poden ser estàtiques (també anomenades d'estat o situació) o dinàmiques (anomenades de transició).
- **Operacions contra la integritat operacional.** En sistemes multiusuari és imprescindible un mecanisme de control de concurrència per a conservar la integritat de la base de dades. Altrament, es podrien produir importants inconsistències derivades de l'accés concurrent.

5. Tasques/responsabilitats de l'administrador

Per tant, amb tot el que hem vist, una relació aproximada de les tasques/responsabilitats de l'administrador de dades és la següent:

- Vetllar perquè les dades siguin als servidors.
- Tenir cura de la còpia de seguretat de les dades i dissenyar, si cal, la política de còpies de seguretat.
- Vetllar perquè els permisos de totes les dades siguin correctes i ningú no pugui accedir a més informació de la que necessita, o hi pugui accedir per protecció de dades.
- Assegurar la disponibilitat de les dades als usuaris.
- Vetllar per un temps de resposta de les bases de dades correcte.
- Saber on hi ha totes les dades/bases de dades de l'organització.
- Evitar al màxim la duplictat d'informació dins l'organització.
- Resoldre les consultes de la direcció al sistema informàtic.
- Configurar bases de dades corporatives.
- Establir els permisos i els accessos a aquestes bases de dades per part dels usuaris.
- Vetllar pel funcionament correcte de les bases de dades.
- Vetllar per una gestió i un emmagatzematge correctes de les dades segons els protocols d'actuació definits en la Llei orgànica de protecció de dades (LOPD).

Vegeu també

Vegeu el mòdul "Administració de la seguretat".

Resum

Les dades són la base del sistema informàtic i també són de gran valor per a l'organització. El gran perill és que poden ser en molts llocs. Hem d'intentar d'educar els usuaris perquè les concentrin en els servidors d'una manera natural. Això dona uniformitat i seguretat a les dades.

Malgrat tot, això no és suficient per a satisfer una de les grans demandes de l'organització: extreure noves conclusions a partir de la informació. L'única manera és concentrant la informació en una única aplicació, un servidor de bases de dades.

Si optem per aquesta solució, hem d'instal·lar una interfície homogènia a sobre d'aquest servidor de bases de dades, una arquitectura de dades, un sistema de planificació de recursos de l'empresa (ERP) o un magatzem de dades.

En haver-se convertit en un actiu molt valuós, la informació cal que estigui protegida; així, són necessàries tant polítiques de còpies de seguretat com proteccions d'accés a les dades.

Activitats

1. Observeu organitzacions del vostre entorn (botigues, empreses, bancs, etc.) i intenteu de veure les fonts de dades que hi pot haver.
2. Fixeu-vos en l'entorn. Veureu la mateixa informació amb molts formats diferents, per exemple, l'hora (digital, analògic, etc.). Com ho guardaríeu en una base de dades? Mireu i trobareu altres exemples de dades amb multitud de formats diferents.

Exercicis d'autoavaluació

1. Enumereu algunes de les fonts de dades dels llocs següents:
 - a) Un hospital.
 - b) Una discoteca.
2. Un dels comercials de l'organització us diu que per a fer els càlculs de venda dels productes, per a guardar els productes i els increments, no vol fer servir una base de dades, perquè mai no ha treballat així: ell negocia amb el client directament i, com a molt, pot introduir les dades en un full de càlcul al seu portàtil (i encara us farà un gran favor, si ho fa). Què li diríeu?
3. Concentraríeu les dades en els servidors?
 - a) Sí, perquè així es poden compartir més fàcilment.
 - b) No, és més pràctic distribuir-les en els llocs en què es necessiten.
 - c) No, perquè podríem col·lapsar els servidors i la xarxa.
 - d) Sí, perquè és més senzill fer consultes en la informació de l'organització.
 - e) Sí, perquè eliminem completament la redundància.
 - f) a i d.
4. Quina d'aquestes frases sobre els servidors de bases de dades és falsa?
 - a) Els servidors de bases de dades concentren la informació dispersa en un sol lloc.
 - b) Els ERP tenen com una de les seves bases un servidor de bases de dades.
 - c) Un servidor de bases de dades amb la informació de l'organització facilita les consultes de la direcció.
 - d) Amb un servidor de bases de dades el trànsit de la xarxa disminueix perquè els formats de sortida són homogenis.
 - e) Un servidor de bases de dades homogeneïtza la manera de guardar la informació.
5. Una d'aquestes tasques no és responsabilitat de l'administrador de dades.
 - a) Evitar al màxim la duplictat d'informació dins l'organització.
 - b) Assegurar la disponibilitat de les dades als usuaris.
 - c) Configurar bases de dades corporatives.
 - d) Connectar les bases de dades amb el servidor web.
 - e) Vetllar pel funcionament correcte de les bases de dades.

Solucionari

Exercicis d'autoavaluació

1.a) Es genera informació en molts llocs. Alguns són:

- Recepció de pacients
- Estendre receptes.
- Visites a totes les consultes
- Tot el departament de comptabilitat
- Tot el departament de facturació a pacients
- Tot el departament de comandes a proveïdors
- Departament de nòmines

b) També es genera informació en molts llocs. Alguns són:

- Màrqueting
- Contractar espectacles
- Comptabilitat
- Facturació
- Nòmines
- Comandes a proveïdors
- Seguretat

2. Si una persona us exposa aquest problema, se l'ha de convèncer amb arguments com, per exemple, els següents:

- L'organització treballa com una unitat i, per tant, tots els preus són en un sol lloc, dins d'una base de dades en un servidor.
- Pot agafar les dades i manipular-les, però si negocia canvis de preu els ha de reflectir en la base de dades per a saber:
 - Quin és el venedor més competitiu.
 - Quin és el venedor que treballa millor.
 - Quin és el venedor que factura més.

D'aquesta manera, l'organització pot assignar incentius (econòmics) o d'un altre tipus als comercials amb millors índexs de venda.

Si només té les dades en local i perd l'ordinador, li roben, etc., ni ell ni ningú de l'organització no les podrà recuperar, per la qual cosa l'organització tindrà una pèrdua d'informació i ell pot perdre els incentius que hem comentat.

La informació de l'organització és propietat de l'organització. Per tant, és l'organització qui estableix les directrius de funcionament, i no les persones que la manipulen.

3. f

4. d

5. d

Glossari

base de dades *f* Terme genèric que indica un lloc per a guardar dades.

dada *f* Registre dels successos.

datawarehouse *m* Base de dades que emmagatzema una gran quantitat de dades transaccionals integrades per a ser usades per a l'anàlisi.

Data Mart *m* Conjunt de fets i dades organitzats per a suport decisional basats en la necessitat d'una àrea o departament específic. Les dades són orientades a satisfer les necessitats particulars d'un departament i tenen sentit només per al personal d'aquest departament.

data mining *m* Anàlisi de les dades per a descobrir relacions, patrons o associacions desconegudes.

diccionari de dades *m* Compendi de definicions i especificacions per a les categories de dades i les seves relacions.

dimensió *f* Entitat independent dins del model multidimensional d'una organització que serveix com a clau de recerca (actuant com a índex) o com a mecanisme de selecció de dades.

DSS *m* Vegeu **sistema de suport de decisions**.

enterprise resource planning *m* Vegeu **planejament de recursos de l'empresa**.

ERP *m* Vegeu **planejament de recursos de l'empresa**.

font de dades *f* Qualsevol element (que pertany o no a l'organització) que crea dades que fa servir el sistema informàtic.

informació *f* Processament de les dades perquè tinguin sentit.

llenguatge de consultes estructurat *m* Llenguatge que es fa servir per a interrogar els gestors de bases de dades.

en structured query language.

sigla: **SQL**.

planejament de recursos de l'empresa *m* Sistema de gestió de la informació integrat que pretén donar una solució completa al problema de la informació dins d'una organització.

en enterprise resource planning.

sigla: **ERP**.

OLAP *m* Sigla d'*on-line analytical processing*. Conjunt de principis que proveeixen un entorn de treball dimensional per a suport decisional.

OLTP *m* Sigla d'*on-line transaction processing*. Sistema transaccional que manté les dades operacionals de l'organització.

servidor de bases de dades *m* Sistema que guarda dades i que, per algun mecanisme, generalment mitjançant peticions a través de la xarxa, se li demana informació o se n'hi introdueix.

SGBD *m* Vegeu **sistema de gestió de bases de dades**.

sistema de gestió de bases de dades *m* Programari que gestiona dades d'una manera ordenada per a guardar-les i recuperar-les.

sigla: **SGBD**.

sistema de suport a la presa de decisions *m* Sistema d'aplicacions automatitzades de l'organització que assisteix en la presa de decisions mitjançant una anàlisi estratègica de la informació històrica.

SQL *m* Vegeu **llenguatge de consultes estructurat**.

structured query language *m* Vegeu **llenguatge de consultes estructurat**.

Bibliografia

Barcelo García, M.; Pastor i Collado, J. (1999). *Gestió d'una organització informàtica*. Barcelona: Universitat Oberta de Catalunya.

Date, C. J. (2000). *An Introduction to Database Systems*. Estats Units: Addison Wesley.

Elmasri, R.; Navathe, S. (2000). *Fundamentals of Database Systems*. Estats Units: Addison Wesley.

Inmon, W. (2005). *Building de Data Warehouse*. Estats Units: Wiley. 4a. ed.

Kimball, R.; Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Estats Units: Wiley. 2a. ed.

Pfleeger, C. (1997). *Security in Computing*. Estats Units: Prentice Hall.

Prague, C.; Irwin, M. (1996). *El libro de Access para Windows 95*. Madrid: Anaya multimedia S.A.