

7 SISTEME DE GESTIUNE A BAZELOR DE DATE DISTRIBUITE

7.1 Concepte

Sistemele de baze de date au evoluat de la stadiul în care fiecare aplicație își întreține propria bază de date, la cel în care datele sunt definite și gestionate centralizat. Acum însă, tehnologia bazelor de date **distribuite** poate schimba modul de lucru centralizat într-unul descentralizat. Tehnologia SGBDD constituie una din realizările majore în domeniul sistemelor de baze de date.

Bază de date distribuită	o colecție <u>partajată</u> de date <u>elementare</u> sau <u>structurate</u> , între care există <u>relații logice</u> , proiectată pentru a satisface nevoile informaționale ale unei organizații și <u>distribuită</u> din punct de vedere fizic într-o <u>rețea de calculatoare</u> .
---------------------------------	--

SGBD distribuit	un sistem de programe care permite <u>gestionarea bazei de date distribuite</u> și face ca <u>distribuirea să fie transparentă</u> pentru utilizatori.
------------------------	--

Un SGBDD este format dintr-**o singură BD logică**, divizată într-un număr de **fragmente**. Fiecare fragment este stocat pe unul sau mai multe calculatoare, sub controlul unui SGBD separat, cu calculatoarele legate printr-o rețea de comunicații. Fiecare *site* este capabil de a prelucra independent cererile utilizatorilor care solicită accesul la datele locale (adică are un oarecare grad de autonomie locală) și, deasemenea, este capabil să prelucreze datele stocate pe alte calculatoare din rețea.

Utilizatorii accesează baza de date prin intermediul aplicațiilor. Aplicațiile se clasifică în:

- **aplicații locale**, care nu necesită date din alte *site*-uri
- **aplicații globale**

Este necesar ca un SGBDD să conțină cel puțin o aplicație globală.

Prin urmare, un SGBDD are următoarele **caracteristici**:

- conține o colecție de date partajate și corelate logic
- datele sunt divizate într-un număr de fragmente
- fragmentele pot fi reproduse
- fragmentele/reproducerile sunt alocate *site*-urilor
- *site*-urile sunt legate într-o rețea de comunicații
- datele din fiecare *site* se află sub controlul unui SGBD
- sistemul SGBD din fiecare *site* poate trata autonom aplicațiile locale
- fiecare sistem SGBD participă la cel puțin o aplicație globală

Nu este necesar ca fiecare *site* din sistem să aibă propria bază de date locală.

Așa cum reiese și din definiție, un SGBDD trebuie să facă **distribuirea transparentă** (invizibilă) pentru utilizator. Astfel, faptul că o BD distribuită este divizată în fragmente care pot fi stocate pe calculatoare diferite și eventual reproduse, trebuie ascuns față de utilizator. **Obiectivul transparenței este de a face ca sistemul distribuit să apară ca un sistem centralizat.** Acesta este denumit uneori **principiul fundamental** al SGBDD.

Prelucrarea distribuită

Este important să se facă deosebirea între un SGBDD și **prelucrarea distribuită**.

Prelucrarea distribuită	presupune o bază de date centralizată care poate fi accesată printr-o rețea de calculatoare.
--------------------------------	--

Punctul cheie în definirea unei **baze de date distribuite** constă în faptul că sistemul este format din **date care sunt distribuite fizic în rețea**. Dacă datele sunt centralizate, chiar dacă alți utilizatori le pot accesa prin rețea, acesta nu este considerat un SGBDD, ci o simplă prelucrare distribuită.

Sisteme SGBD paralele

Deasemenea trebuie făcută distincție între un SGBDD și un **SGBD paralel**.

SGBD paralel	este un SGBD rulat pe mai multe procesoare și discuri , care au fost proiectate să execute operațiile în paralel pentru a îmbunătăți performanțele.
---------------------	---

Sistemele SGBD paralele se bazează pe premiza că sistemele cu un singur procesor nu mai pot îndeplini cerințele crescânde privind scalabilitatea eficientă ca fiabilitate, performanțe și cost. În sistemele SGBD paralele se leagă **mai multe mașini mai mici** pentru a obține același transfer ca al unei **mașini mai mari**.

Pentru a asigura accesul mai multor procesoare la o singură BD, sistemul SGBD paralel trebuie să facă posibilă administrarea partajată a resurselor. Resursele partajate și modul cum au fost implementate afectează în mod direct performanțele și scalabilitatea sistemului. Sunt 3 arhitecturi principale pentru sistemele SGBD paralele:

- memorie partajată
- disc partajat
- nimic partajat

Memoria partajată reprezintă o arhitectură *strâns cuplată*, în care mai multe procesoare din cadrul unui **singur sistem** partajează memoria acestuia. Cunoscută sub numele de **prelucrare simetrică multiplă (SMP)** această tratare este larg utilizată pe diverse platforme, de la stații de lucru personale care acceptă câteva microprocesoare în paralel, până la marile mașini RISC și chiar cele mai mari calculatoare mainframe. Această arhitectură oferă un acces foarte rapid la date pentru un număr limitat de procesoare, dar nu este scalabil la mai mult de 64 procesoare, moment în care rețeaua de interconectare se strangulează.

Arhitectura cu **nimic partajat**, cunoscută și sub numele de **prelucrare paralelă masivă (MPP)**, este o arhitectură cu procesoare multiple, în care **fiecare face parte dintr-un sistem complet**, cu propria sa memorie și capacitate de stocare pe disc. BD este partiționată între toate discurile sistemelor asociate acesteia, iar datele sunt disponibile transparent pentru utilizatorii din toate sistemele. Această arhitectură este mai scalabilă decât cea cu memorie partajată și poate accepta cu ușurință un număr mare de procesoare. Totuși, performanțe bune se obțin când datele cerute sunt stocate local.

Discul partajat reprezintă o arhitectură *cuplată lejer*, optimizată pentru aplicații care sunt inerent centralizate și necesită un înalt grad de disponibilitate și performanță. Fiecare procesor poate accesa direct toate discurile, dar fiecare are propria sa memorie. Ca și arhitectura cu nimic partajat, aceasta elimină strangularea performanțelor datorită partajării memoriei.

De obicei, tehnologia paralelă este utilizată pentru **BD foarte mari**, de ordinul teraoctetilor sau pentru sistemele care trebuie să prelucreze mii de tranzacții pe secundă.

7.2 Avantajele și dezavantajele SGBDD

Distribuirea datelor și aplicațiilor prezintă o serie de avantaje potențiale față de sistemele centralizate tradiționale.

Avantaje

1. **structura organizațională:** adecvată pentru organizații cu filiale în diferite localități
2. **caracterul partajabil și autonomia locală**
3. **disponibilitate crescută:** o pană în cadrul unui *site* sau o pană a unei linii de comunicație nu face ca întregul sistem să devină inoperabil
4. **fiabilitate crescută:** datorită posibilității reproducerii datelor în cadrul mai multor *site-uri*, pana unui nod sau a unei linii de comunicație nu implică neapărat ca datele să fie inaccesibile
5. **performanțe îmbunătățite:** comparativ cu o bază de date situată la distanță, viteza de acces la baza de date distribuită este mai bună, întrucât datele cel mai des accesate sunt localizate în apropierea *site-ului* cu cea mai mare cerere.
6. **economie:** este mai eficientă adăugarea de noi stații de lucru într-o rețea decât să se modernizeze un sistem mainframe
7. **dezvoltare modulară:** într-un mediu distribuit este mult mai ușor să se trateze expansiunea: pot fi adăugate *site-uri* noi, creșterea dimensiunii bazei de date poate fi rezolvată adăugând în rețea putere de calcul și de stocare

Dezavantaje

1. **complexitate**
2. **cost:** există costuri continue de comunicații, datorate utilizării rețelei și costuri de personal care administrează și întreține sistemele SGBD locale
3. **securitate**
4. **control dificil al integrității**
5. lipsa de **standarde**
6. lipsa de experiență: la ora actuală se află în uz numai câteva SGBDD prototip și specializate
7. proiectarea mai complexă a BD

7.3 Sisteme SGBDD omogene și heterogene

Într-un **sistem omogen**, toate *site-urile* utilizează același SGBD. Sistemele omogene sunt mai ușor de proiectat și gestionat, adăugarea unui nou *site* în SGBDD este mai ușoară și se obțin performanțe crescute prin exploatarea capacității de prelucrare paralelă a mai multor *site-uri*.

Într-un **sistem heterogen**, *site-urile* pot utiliza SGBD diferite, bazate pe modele de date diferite: relațional, în rețea, ierarhice, orientate spre obiecte. De obicei sistemele heterogene se creează atunci când *site-urile* individuale și-au implementat propriile baze de date, iar integrarea este avută în vedere într-o etapă ulterioară. Într-un sistem heterogen sunt necesare traduceri pentru a permite comunicarea dintre diversele SGBD. Pentru a realiza transparența sistemului, este necesar ca utilizatorii să poată efectua cererile în limbajul sistemului SGBD din *site-ul* local, apoi sistemul are sarcina de a localiza datele și de a efectua orice traducere necesară.

Soluția tipică într-un **SGBDD heterogen** constă în utilizarea unor **porți** care transformă limbajul și modelul fiecărui SGBD în limbajul și modelul relațional. Totuși, tratarea prin porți are și o serie de limitări serioase legate de **administrarea tranzacțiilor** și **controlul concurenței**.