

Cap.1. Sisteme distribuite

1.1. Sistemul distribuit. Concepte, definiții

Centralizare și descentralizare – două tendințe opuse – au marcat alternativ evoluția sistemelor informatice. Sistemele distribuite reprezintă unul dintre compromisurile cele mai profitabile, păstrând avantajele ambelor abordări. Trecerea de la centralizare la descentralizare a fost marcată în primul rând de reducerea costurilor de comunicație, pe de o parte, iar pe de altă parte autonomia locală a reprezentat unul dintre cele mai importante avantaje oferite de sistemele distribuite.

Tratarea sistemelor distribuite necesită în primul rând o definiție clară și precisă a conceptelor. Semnificația cuvântului *distribuit* din DEX este cea de *împărțit* sau *repartizat*. Aceasta reflectă că împărțirea, divizarea sau repartitia se realizează conform unor reguli, norme de clasificare. De aici rezultă că un sistem distribuit, nu neapărat informatic, este un ansamblu de entități (fragmente) eterogene, repartizate la noduri diferite, care conlucrează în vederea realizării unor obiective, fiecare dintre ele având un anumit nivel ridicat de autonomie.

Definiția noțiunii de sistem distribuit este prezentată în literatura de specialitate sub diverse forme.

În *Enciclopedia de Computer Science* [WILEY04] definiția sistemelor distribuite se face printr-o paralelă cu distribuția geografică a calculatoarelor în locații diferite unele de altele. Ca exemplu este prezentat un sistem enterprise business care trebuie să suporte utilizatori diferiți din zone diferite, de pe site-uri web diferite. *Sistemul distribuit se definește "ca un ansamblu de aplicații software, hardware-ul pe care acestea rulează și legăturile de comunicații dintre dispozitivele hardware"* [WILEY04]. Internetul este cel mai mare și cel mai cunoscut sistem distribuit.

În dicționarele și enciclopediile anglo-saxone, pentru *distributed* există următoarele două semnificații importante: caracterizat printr-o distribuție statistică de un anumit fel (ex. distribuție normală) și respectiv rețea de calculatoare în care cel puțin o parte din procesare este realizată de stațiile de lucru autonome fiind împărțită și memorată de calculatoarele din rețea.

Din punct de vedere matematic, cuvântul *distributed* sau *distributive* semnifică distributivitatea înmulțirii față de adunare [ECICLOPEDIA BRIT]

În literatura anglo-saxonă, este folosită sintagma *distributed system* cu semnificația de sistem distribuit, explicația terminologiei fiind în strânsă legătură cu etimologia sa. În Encyclopaedia Britannica, verbul *dis-trib-ute*, cu originea în latinescul *distributus*, (cuvânt compus din prefixul *dis-* + și verbul *tribuere*) are sensul este tot de a împărți (*to divide into several or many*). Acest lucru atrage atenția către următoarea semnificație din Encyclopaedia Britannica, *to spread out so as to cover something* care s-ar traduce a *împrăștiia cu scopul de a acoperi ceva*. Cea din urmă semnificație consider că este cea mai importantă în privința domeniului pe care-l abordez în aceasta lucrare. A distribui mai semnifică și împărțirea pe categorii, relevantă pentru scopul lucrării.

Trăsăturile definitorii ale sistemelor distribuite, așa cum sunt prezentate în [WILEY04] sunt următoarele: noduri multiple, partajarea resurselor, concurență, transmiterea de mesaje (*message passing*), eterogenitatea nodurilor, protocoale multiple, deschidere, transparența defectelor / toleranța la defecte, persistența, securitatea, autonomia locală, inexistența unui server central.

Nodurile multiple semnifică existența mai multor calculatoare (noduri) legate între ele în rețea. Este esențială distincția între procesarea paralelă care presupune existența a mai multe microprocesoare pe același calculator și procesarea distribuită, în care fiecare nod poate lua atât rolul clientului, când solicită servicii de la alt nod, cât și rolul serverului, când oferă servicii altor noduri.

Partajarea resurselor hardware, software precum baze de date, servicii de mail, sau aplicații colaborative este motivul conectării mai multor calculatoare în rețea și lucrul în mod distribuit.

Concurența presupune ca aplicațiile de la mai multe noduri să utilizeze simultan același set de funcționalități independente distribuite în rețea.

Transmiterea de mesaje (message passing) se realizează cu ajutorul protocoalelor de comunicare, cum ar fi TCP/IP având ca suport tehnologii de conectare (Ethernet, modem).

Eterogenitatea nodurilor din punct de vedere hardware și software dă flexibilitate sistemelor, eventualele neajunsuri sau incompatibilități fiind rezolvate prin formate de mesaje agreate și prin protocoale. *Protocoalele multiple* sunt asincrone de cele mai multe ori.

Deschiderea sistemelor distribuite dă posibilitatea de a adăuga noduri și aplicații noi cât timp sistemul funcționează, spre deosebire de programele secvențiale comune care sunt închise și nu își schimbă configurația în timpul execuției. Fiecare nod al sistemului distribuit îndeplinește un set de condiții și protocoale minimale care asigură interoperabilitatea cu componentele adăugate/modificate în sistem.

Transparența defectelor / toleranța la defecte dă posibilitatea sistemului distribuit să funcționeze chiar dacă există componente ale care sunt nefuncționale, lucru care nu trebuie să afecteze funcțiunile de ansamblu ale sistemului și nici nu trebuie să fie observate de utilizator.

Persistența presupune existența de date și informații stocate în medii persistente, nevolatile iar *securitatea* asigură accesul fiecărui utilizator la informațiile corespunzătoare drepturilor pe care acesta le are.

Autonomia locală a fiecărui nod este fizică și logică, astfel el comunică cu celelalte noduri printr-un set de protocoale bine definite. *Inexistența unui server central* presupune ca funcționalitatea sistemului să fie asigurată de domenii de agenți autonomi care veghează asupra configurației, securității, integrității și managementului sistemului.

În lucrarea [TOMO08], un sistem distribuit este definit ca “o colecție de programe și procese care operează peste o rețea de echipamente electronice, în care fiecare nod are autonomie locală, propria unitate de procesare și memorie iar nodurile în ansamblul lor sunt transparente pentru utilizator în ideea realizării obiectivului urmărit”.

Analizând definițiile prezentate rezultă că un sistem distribuit cuprinde trei niveluri de distribuire: de echipamente, de aplicații informatice și de date. În concluzie, consider că un sistem distribuit poate fi definit ca un ansamblu de echipamente hardware legate în rețea, o infrastructură de comunicații între nodurile autonome, date și aplicații software distribuite pe nodurile rețelei de calculatoare care permit partajarea unui set minim de resurse la costuri mici.

1.2. Avantajele și dezavantajele sistemelor distribuite

Distribuirea datelor și aplicațiilor prezintă potențiale avantaje și respectiv dezavantaje față de centralizare.

1.2.1. Avantajele sistemelor distribuite

- *Structura organizațională.* Unele companii își desfășoară activitatea în mai multe locații geografice, situate la distanțe importante unele de celelalte. În mod natural, un sistem informatic trebuie să respecte fluxul informațional din cadrul organizației. Astfel, datele consultate la o filială a organizației sunt, cu preponderență datele locale și doar sporadic, date globale ale companiei. Deci este mult mai simplu ca locația respectivă să păstreze și să administreze doar acea parte de date pe care o folosește în mod curent.
- *Caracterul partajabil și autonomia locală.* Utilizatorii din cadrul unui nod pot accesa datele din alte noduri. Datele pot fi plasate în nodul cel mai apropiat de utilizatorii care le folosesc în mod normal. Astfel, utilizatorii dețin controlul local asupra datelor, putându-se stabili și impune politici locale referitoare la utilizarea acestora. Există un administrator al bazei de date globale care este responsabil de toată baza de date. El transferă o parte din responsabilități administratorilor locali ai bazei de date.
- *Disponibilitate crescută.* Defectarea calculatorului sau a liniilor de comunicație către un SGBD centralizat duce la întreruperea completă a lucrului, spre deosebire de un SGBD distribuit unde defectarea calculatorului la un nod sau întreruperea comunicației cu acel nod nu face ca întreg sistemul să fie inoperabil, iar cererile îndreptate către nodul defect se redirecționează către alte noduri din sistem.
- *Fiabilitate crescută.* Datorită replicării datelor la mai multe noduri, în cazul unei defecțiuni a unui nod, datele aferente lui se regăsesc și pe alte noduri, ele fiind în continuare accesibile.
- *Performanțe îmbunătățite.* Deoarece la fiecare nod volumul datelor stocate și numărul de accesări ale datelor este mai mic decât ar fi dacă am avea un sistem centralizat aflat la distanță, timpul de răspuns este mai mic. Pe de altă parte, datele de care au nevoie utilizatorii de la un nod sunt stocate la nodurile cele mai apropiate, comunicația ar putea fi mai rapidă.
- *Economie.* Din punct de vedere al costului, este mult mai ieftin să realizezi o rețea de calculatoare mici, cu putere echivalentă a unui calculator de mare putere. Un alt aspect îl reprezintă extinderea sistemului care este mai facilă prin adăugarea de calculatoare mici într-o rețea decât prin modernizarea unui

mainframe. Se poate vorbi și de o economisire a costului comunicației, acolo unde bazele de date sunt îndepărtate din punct de vedere geografic și aplicațiile necesită accesul la date aflate la distanță. Ar fi o soluție mai ieftină partiționarea bazei de date și prelucrarea locală decât prelucrarea la distanță.

- *Dezvoltare modulară.* Este, în mod firesc, mai ușor de tratat extinderea unui sistem într-un mediu distribuit, unde pot fi adăugate noduri noi relativ ușor. Mărirea dimensiunii bazei de date poate fi tratată prin includerea în rețea putere de calcul și de stocare. Într-un sistem centralizat această problemă ar putea conduce la necesitatea schimbării mainframe-ului și chiar și a SGBD-ului pentru a satisface noile cerințe ale organizației.

1.2.2. Dezavantajele sistemelor distribuite

- *Complexitatea.* Într-un sistem distribuit, în care natura sa distribuită este ascunsă față de utilizator și care produce un nivel acceptabil al performanțelor – fiabilitatea și disponibilitatea sunt inerent mai complexe decât într-un sistem SGBD centralizat. De asemenea, faptul că datele pot fi reproduse adaugă sistemului SGBD un nivel suplimentar de complexitate.

- *Costul.* Complexitatea crescută înseamnă că este de așteptat ca prețurile procurării și întreținerii unui sistem distribuit să fie mai mari decât ale unuia centralizat. În plus, mai apar și costurile de comunicație, precum și costurile cu personalul care administrează și întreține sistemele SGBD locale și rețeaua acestora.

- *Securitatea.* Pe lângă securizarea accesului la datele de la noduri, trebuie securizată și rețeaua de comunicație.

- *Controlul integrității este mai dificil* deoarece necesită accesul la o cantitate mare de date și duce la mărirea timpului comunicațiilor și crește complexitatea prelucrărilor necesare pentru impunerea constrângerilor de integritate.

- *Lipsa de standarde* care limitează substanțial potențialul SGBD distribuite. De asemenea, nu există instrumente sau metodologii care să ajute utilizatorii să transforme un sistem SGBD centralizat într-unul distribuit.

- *Lipsa de experiență* care face anevoioasă dezvoltarea și implementarea.

- *Proiectarea mai complexă a bazei de date.* În afară de dificultățile normale ale proiectării unei baze de date centralizate,

în cazul variantei distribuite trebuie luate în considerare fragmentarea datelor, alocarea fragmentelor unor anumite noduri și replicarea datelor.

1.3. Sisteme de baze de date distribuite

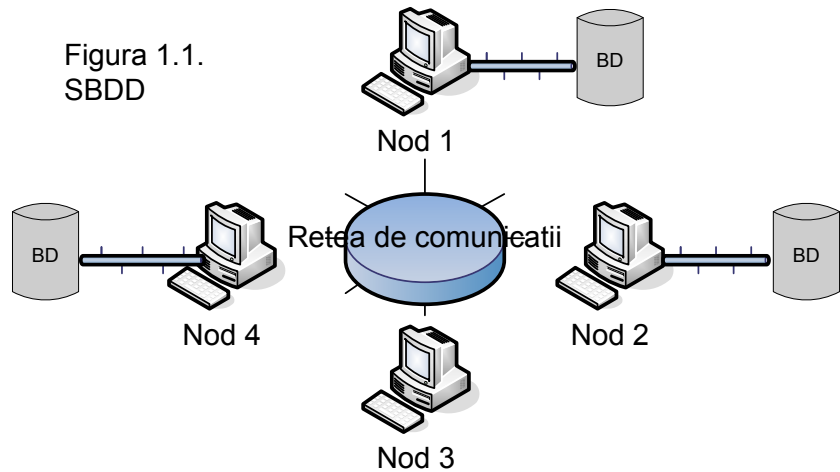
1.3.1. Definirea sistemelor de baze de date distribuite

Conform lucrării [TOMO08], *sistemele de baze de date distribuite* reprezintă colecții de noduri conectate în rețea, fiecare dintre noduri fiind o bază de date care comunică și colaborează cu celelalte baze de date astfel încât utilizatorul să poată accesa date de oriunde în interiorul rețelei, într-un mod transparent.

Sistemele de baze de date distribuite sunt compuse în principal din bazele de date distribuite și sistemul de gestiune a bazelor de date distribuite.

Sistemele de baze de date distribuite (SBDD) trebuie să ajute la rezolvarea problemei “insulelor informaționale”, apărute ca urmare a separării geografice, a arhitecturilor incompatibile ale calculatoarelor, a protocoalelor de comunicație incompatibile, și așa mai departe. Integrarea sistemelor de baze de date într-un întreg din punct de vedere logic ar putea opri acest mod de gândire.

Așa cum se observă din fig.1.1, nu este obligatoriu ca fiecare nod să aibă propria bază de date.



Sistemul de baze de date distribuite cuprinde următoarele **componente software**: *componenta de date, componenta hardware și de comunicație, sistemul de gestiune a bazelor de date locale, dicționar de date global și sistemul de gestiune al bazei de date distribuite.*

Componenta de date cuprinde baza de date distribuită.

Componenta de hardware și comunicație cuprinde rețeaua de calculatoare și infrastructura de comunicații care realizează legăturile în cadrul rețelei. Ea cuprinde descrierea completă a nodurilor și a legăturilor din cadrul rețelei.

Sistemul de gestiune al bazei de date locale împreună cu baza de date locală este un sistem standard de gestiune a bazelor de date. El are propriul dicționar pentru datele locale.

Dicționarul de date global cuprinde informații despre baza de date distribuită: localizare, structură, disponibilitate și modul de utilizare a datelor.

Sistemele de gestiune a bazelor de date distribuite (DDBMS) constituie software-ul prin care mai multe baze de date locale sunt legate printr-un sistem de comunicație astfel încât datele din orice stație să poată fi accesate de utilizatorii de la alte stații. Altfel spus, el cuprinde un sistem complex de programe care asigură interfața între baza de date distribuită și utilizatorii acesteia.

1.3.2. Analiza obiectivelor bazelor de date distribuite

O *bază de date distribuită* este o bază de date controlată de un sistem de gestiune a bazelor de date central și în care mediile de stocare a datelor se regăsesc la nodurile unei rețele de calculatoare. Baza de date este fragmentată în partiții aflate în locații diferite, fiecare partiție/fragment putând fi replicat sau nu.

Adesea se face confuzie între o bază de date distribuită și o bază de date cu prelucrare distribuită. *Prelucrarea distribuită* reprezintă accesarea unei baze de date centralizate prin intermediul unei rețele de calculatoare. O *bază de date distribuită*, conform [OZVA91], reprezintă o colecție de date intercorelate logic și distribuite fizic într-o rețea de calculatoare.

În definiția [OZVA91] sunt punctate două aspecte importante ale bazelor de date distribuite. În primul rând, există o singură bază de date din punct de vedere logic, iar utilizatorul interacționează cu aceasta în mod similar cu bazele de date centralizate. Autorii scot astfel în evidență că datele sunt integrate logic. În al doilea rând, autorii subliniază existența unei distribuiri fizice a datelor cu care se lucrează, așadar există o partiționare fizică a bazei de date pe stațiile unei rețele de calculatoare. Fiecare din mașinile de la nodurile rețelei are autonomie de prelucrare, astfel stația locală participă la execuția aplicațiilor globale care necesită accesarea datelor din mai multe stații, dar, în același timp, ea poate să realizeze aplicații locale.

Obiectivele bazelor de date distribuite, în mod similar bazelor de date tradiționale se referă la *controlul centralizat, independența datelor, redundanța minimă și controlată a datelor, accesul eficient la date, integritatea și securitatea datelor.*

În cazul bazelor de date distribuite **controlul centralizat este mai accentuat decât în cazul bazelor de date centralizate. La acestea se identifică o structură de control ierarhic bazată pe un administrator global care are responsabilitatea centrală a întregii baze de date distribuite și pe administratorii locali care pot avea un înalt grad de autonomie care poate merge până la realizarea coordonării între stații. Bazele de date distribuite diferă foarte mult în privința autonomiei stațiilor, aceasta putând varia de la autonomia completă, fără existența administratorului global, până la cel mai complet control centralizat.**