

## PARTIȚIONAREA BAZELOR DE DATE

Victor ȘAPTEFRAȚI

Departamentul Ingineria Software și Automatică, grupa TI-191 F/R, Facultatea CIM, Universitatea Tehnică a Moldovei, Chișinău, Republica Moldova

Autorul corespondent: Victor Șaptefrați, e-mail: [saptefrativictor@gmail.com](mailto:saptefrativictor@gmail.com)

Conducător științific: Dorian SARANCIUC, DISA, FCIM, UTM

**Rezumat.** Conceptul de partiționare este o tehnică de descompunere a bazelor de date care o împarte în unități mai mici, independente, numite partiții (sau „shards”). Scopul este de a distribui datele și sarcina pentru a îmbunătăți performanța și scalabilitatea sistemului de baze de date, precum și disponibilitatea datelor în caz de accidente. Punerea în aplicare necesită o planificare atentă, inclusiv alegerea cheii de partiționare, arhitectura partiționării, localizarea fragmentelor și gestionarea comunicării între fragmente.

**Cuvinte cheie:** Partiționare, scalabilitate, distribuție, cheie de partiție.

### Introducere

Partiționarea bazelor de date este o tehnică care permite scalarea bazelor de date relaționale prin distribuirea seturilor de date mari și a sarcinilor de lucru pe mai multe servere. O soluție promițătoare pentru a face față activităților de producere și de procesare a cantităților mari de date, partiționarea bazelor de date este o tehnică care optimizează stocarea datelor și performanța interogărilor prin împărțirea seturilor de date în segmente mai mici, mai ușor de gestionat [1]. Partiționarea bazelor de date este o abordare eficientă care permite organizațiilor să îmbunătățească performanța, scalabilitatea și mentenanța sistemelor lor de gestionare a datelor. Prin împărțirea unui set mare de date în partiții mai mici, această tehnică permite o alocare mai eficientă a spațiului de stocare, o execuție mai rapidă a interogărilor și îmbunătățirea capacităților de procesare paralelă. Partiționarea poate fi implementată la diferite niveluri, cum ar fi partiționarea orizontală (*sharding*) și verticală, fiecare cu setul său unic de avantaje și cazuri de utilizare.

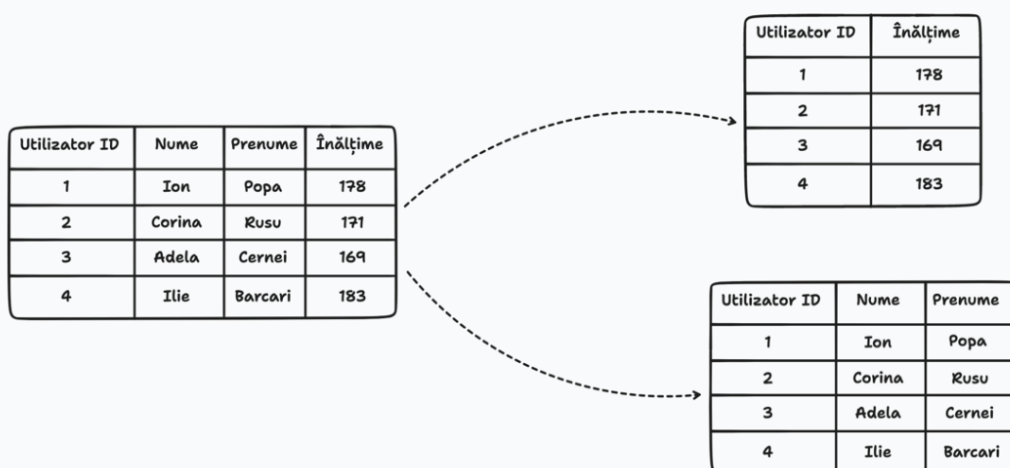


Figura 1. Exemplu de partiționare verticală

### Partiționarea verticală

Cunoscută și sub denumirea de partiționarea pe rânduri, partiționarea verticală este o tehnică care facilitează împărțirea unui tabel mare de bază de date în subansambluri mai mici, mai ușor de gestionat, pe baza valorilor unei coloane cheie de împărțire. Această abordare poate îmbunătăți performanța interogărilor prin reducerea volumului de date care trebuie căutate în timpul operațiunilor de recuperare. Cu toate acestea, ea necesită, de asemenea, punerea în aplicare a unor procese de gestionare suplimentare pentru tabelele mai mici și poate prezenta provocări în ceea ce privește asigurarea coerenței datelor în întregul sistem [2].

În figura 1 se observă conceptul de partiționare verticală, care separă datele coloanelor, *Nume* și *Prenume* și *Înălțime* în două tabele distincte.

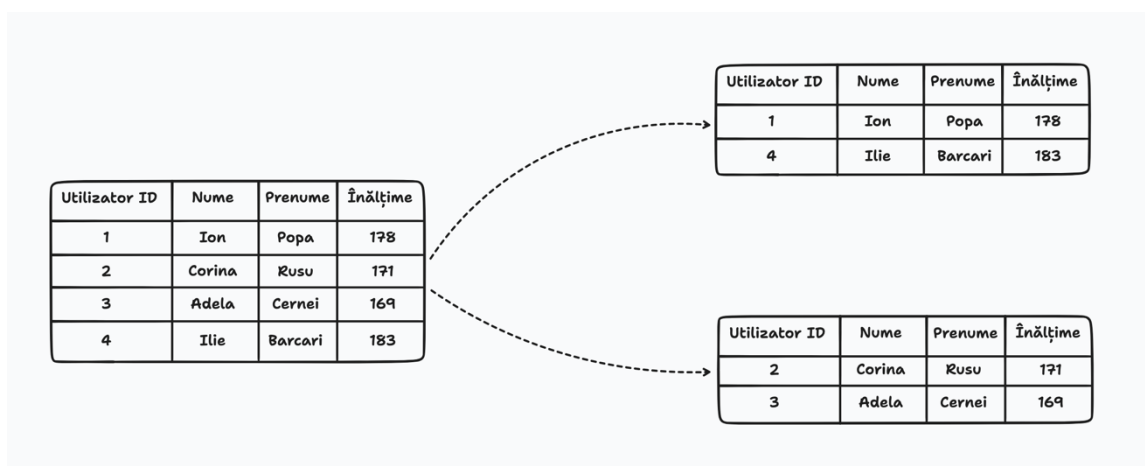


Figura 2. Exemplu de partiționare orizontală

### Partiționarea orizontală

O altă metodă de partiționare este cea orizontală, ilustrată în figura 2, constă în descompunerea tabelului de baze de date în tabele mai mici, în funcție de un anumit interval al cheii de partiție. Acest lucru se realizează prin împărțirea datelor în intervale bazate pe o coloană, cum ar fi data sau ID-ul, și prin crearea unor tabele separate pentru fiecare interval. Acest lucru îmbunătățește performanța și scalabilitatea prin distribuirea datelor și a volumului de lucru pe mai multe servere sau locații și îmbunătățește disponibilitatea datelor și recuperarea în caz de dezastru, permițând realizarea de copii de rezervă și restaurare independente pentru fiecare partiție.

### Beneficiile partiționării

Principalele beneficii ale partiționării unei baze de date sunt:

- *Flexibilitate crescută* – permite scalarea orizontală, prin adăugarea mai multor mașini la o stivă existentă, permițând un trafic mai mare și o procesare mai rapidă.
- *Timp de răspuns mai rapid* – împărțirea unei tabele în mai multe reduce numărul de rânduri pe care o interogare trebuie să le caute, ceea ce duce la timpi de răspuns mult mai rapizi.
- *Fiabilitate crescută* - partiționarea unei baze de date atenuează impactul întreruperilor, minimizând impactul general asupra aplicației sau site-ului web.
- *Gestiunea bazelor de date mari* – partiționarea este deosebit de util pentru aplicațiile cu o bază de date mare, monolitică, care altfel ar fi limitată de capacitatea de stocare și de calcul a unei singure mașini.
- *Securitate îmbunătățită* – partiționarea unei baze de date permite ca datele să fie partiționate și stocate în locații separate, ceea ce face mai dificil pentru hackeri să acceseze informații sensibile.
- *Performanță îmbunătățită* – partiționarea permite ca baza de date să fie distribuită pe mai multe mașini, rezultând o performanță îmbunătățită și o mai bună utilizare a resurselor.

### **Dezavantajele partiționării**

Împărțirea unei baze de date poate îmbunătăți performanța și poate facilita extinderea, dar are și limitări. Una dintre principalele provocări ale partiționării este complexitatea implementării corecte, care poate duce la pierderea de date sau la tabele corupte. O altă problemă este dezechilibrarea partițiilor, care poate cauza încetinirea și blocarea aplicației pentru o parte semnificativă a utilizatorilor. Un alt dezavantaj major al partiționării este faptul că poate fi dificil să readucă baza de date la arhitectura sa originală. Ar trebui să se acorde o atenție deosebită înainte de a decide să se utilizeze o bază de date în mod fragmentat, deoarece este posibil ca aceasta să nu fie cea mai bună soluție pentru orice situație.

### **Arhitectura partiționării**

Atunci când este descompusă o bază de date pe orizontală, este important să se determine modul în care datele vor fi distribuite între fragmente. Există mai multe metode de partiționare, fiecare având avantaje și dezavantaje specifice. Alegerea metodei potrivite depinde de cerințele specifice ale aplicației și de capacitatea de a gestiona și optimiza funcțiile de partiționare, intervalele de valori sau structurile de directoare. Deosebim trei tipuri principale de arhitectură de partiționare:

- **Partiționare bazată pe chei** utilizează o valoare din datele primite, cum ar fi ID-ul unui client, pentru a determina în ce partiție ar trebui să ajungă datele. Aceasta facilitează operațiunile paralele de citire și scriere, deoarece datele pot fi accesate și procesate pe mai multe servere simultan. O funcție partiție este utilizată pentru a converti valoarea într-un ID de partiție. Cu toate acestea, această metodă poate îngreuna adăugarea sau eliminarea dinamică a serverelor în baza de date, de aceea optimizarea funcției de partiționare și implementarea unor strategii eficiente de gestionare a modificărilor pot oferi o soluție robustă și de înaltă performanță pentru aplicațiile moderne.
- **Partiționare bazată pe intervale** presupune împărțirea datelor pe baza unor intervale de valori date, cum ar fi intervalele de prețuri. Această metodă este simplu de implementat, însă aceasta vine cu un dezavantaj prin faptul că nu asigură distribuția uniformă a datelor, selectarea necorespunzătoare a intervalelor de valori poate duce la o încărcare neuniformă a serverului și poate avea un impact asupra performanței. O soluție pentru aceste provocări reprezintă o selecție atentă a intervalelor de valori, ajustări dinamice, monitorizare constantă a performanței și a zonelor cu accesări frecvente ca strategii de optimizare și echilibrare.
- **Partiționare pe bază de directoare** utilizează un tabel de căutare pentru a ține evidența ce date deține o anumită partiție. Această metodă necesită întreținerea tabelului de căutare, dar permite o distribuție mai flexibilă și mai eficientă a datelor și realocarea dinamică a datelor. Oferind cea mai mare flexibilitate, deoarece nu necesită o funcție de partiționare fixă sau intervale predefinite de valori, optimizarea structurii și gestionării tabelului de căutare este importantă și devine o sarcină complexă suplimentară în gestionarea bazei de date. Introducerea tehnicilor precum replicarea tabelului și algoritmi de căutare eficienți sunt esențiali pentru a asigura performanța.

### **Concluzii**

În concluzie, partiționarea bazelor de date reprezintă o soluție eficientă pentru gestionarea și scalarea bazelor de date relaționale, ce vine ca un răspuns provocărilor asociate cu manipularea unor volume mari de date și sarcini de lucru complexe. Prin implementarea tehnicilor de partiționare verticală și orizontală, poate fi îmbunătățită performanța, scalabilitatea și disponibilitatea sistemelor lor de gestionare a datelor. Partiționarea verticală permite o gestionare mai ușoară a tabelelor mari prin divizarea acestora în subsansamble bazate pe coloane-cheie, în timp ce partiționarea orizontală facilitează distribuția datelor și a sarcinilor de lucru pe mai multe servere, îmbunătățind astfel capacitatea de procesare paralelă și redundanța datelor.

Implementarea partiționării bazelor de date necesită o planificare atentă și o gestionare corespunzătoare a cheilor de partiție, arhitecturii și comunicării între fragmente. Provocări precum mentenanța tabelor mai mici și asigurarea coerenței datelor pot apărea în timpul implementării, necesitând strategii adecvate pentru a le soluționa.

#### **Referințe**

1. ERIC RIES, *Sharding for startups* [online]. 2009. [accesat 20.11.2022].  
Disponibil: <http://www.startuplessonslearned.com/2009/01/sharding-for-startups.html>
2. MAHDI YUSUF, *Database Sharding Explained* [online]. 2022 [accesat 22.11.2022].  
Disponibil: <https://architecturenotes.co/database-sharding-explained>