

Partiționarea datelor unei aplicații de tip Cloud pe verticală

Diana Botezatu, gr. 1.2

Tema 2 DATC, 2018-2019

Strategiile de partiționare a datelor sunt foarte importante în cazul aplicațiilor la scară largă, adresând probleme create de volumul mare de date, de rapiditatea cu care volumul de date crește, sau de varietatea de date. Printre cele mai importante avantaje oferite de partiționarea datelor se numără scalabilitatea îmbunătățită având datele împărțite între partiții hardware multiple, un grad de securitate ridicat prin separarea datelor cu caracter secret/sensibil de restul datelor, dar și performanțe crescute.

Unul dintre cele 3 tipuri principale de partiționare este partiționarea datelor pe verticală. Partiționarea pe verticală presupune împărțirea unui tabel în mai multe tabele selectând anumite coloane. Motivul pentru care se folosește această tehnică este îmbunătățirea performanțelor mediilor în care se execută multe operații de citire de date.

În primul rând o astfel de împărțire facilitează citirea datelor de care este nevoie frecvent. Datele accesate des pot fi stocate separat, în tabele mai mici, în timp ce restul datelor pot fi stocate în alte tabele. Exemplu: O aplicație folosește o bază de date stocată în Cloud. Baza de date conține informații despre mai mult echipamente audio. Aplicația utilizează des informațiile din coloanele ID, Produs și Preț pentru a le afișa clienților săi. Din acest motiv, cele 3 coloane au fost separate de restul coloanelor, păstrându-se în ambele tabele id-ul pentru a putea relaționa datele când este cazul.




ID	Produs	Tip	Construcție	Preț	Cantitate
1	Sennheiser RS195	Over-ear	Closed-back	1449 lei	5
2	Sennheiser PXC 250 II	On-ear	Closed-back	620 lei	12
3	Sennheiser Momentum 2.0 On-Ear	On-ear	Closed-back	893 lei	7
4	Sennheiser HD 800 S	Over-ear	Open-back	7199 lei	1
5	Sennheiser HD 201	Over-ear	Closed-back	119 lei	26

ID	Produs	Preț
1	Sennheiser RS195	1449 lei
2	Sennheiser PXC 250 II	620 lei
3	Sennheiser Momentum 2.0 On-Ear	893 lei
4	Sennheiser HD 800 S	7199 lei
5	Sennheiser HD 201	119 lei


ID	Tip	Construcție	Cantitate
1	Over-ear	Closed-back	5

2	On-ear	Closed-back	12
3	On-ear	Closed-back	7
4	Over-ear	Open-back	1
5	Over-ear	Closed-back	26

În al doilea rând o astfel de împărțire poate salva spațiu în Cloud când avem de a face cu baze de date care conțin date binare, de exemplu șiruri de biți provenite din imagini. String-urile asociate înregistrărilor pot fi în continuare salvate în Cloud, în timp ce șirurile de biți pot fi stocate cu ajutorul unor servicii de Blob storage (binary large object), care oferă, la un preț mai scăzut decât prețul scalarării unei baze de date pentru foarte multe intrări de dimensiuni mari, back-up, disponibilitate și securitate. Calea de la fișier va fi în continuare disponibilă în baza de date. Exemplu:

ID	Nume și prenume	Anul nașterii	Imagine
1	Freddie Mercury	1946	
2	James LaBrie	1963	
3	Dave Grohl	1969	

ID	Nume și prenume	Anul nașterii	Path
1	Freddie Mercury	1946	www.storage.com/pathdfg346t46a43
2	James LaBrie	1963	www.storage.com/path3o53dfg84bg0
3	Dave Grohl	1969	www.storage.com/path34tnf98465w0

ID	Imagine
1	
2	
3	

În al treilea rând se pot separa informațiile secrete, iar bazele de date în care se află astfel de informații pot fi securizate suplimentar. Exemplu:

ID	Denumire	Preț	Ingredientul secret
1	Supă cremă de ciuperci	15.99	Cimbru
2	Pui cu parmezan la cuptor	25.99	Usturoi
3	Lasagna	24.99	Sos Bolognese

ID	Denumire	Preț
----	----------	------

1	Supă cremă de ciuperci	15.99
2	Pui cu parmezan la cuptor	25.99
3	Lasagna	24.99

ID	Ingredientul secret
1	Cimbru
2	Usturoi
3	Sos Bolognese

Partiționarea verticală are însă și un dezavantaj și anume nu este potrivită bazelor de date care sunt modificate des deoarece modificarea presupune schimbări în mai multe baze.