



Tema 2 DATC

Partiționarea datelor pentru o aplicație de tip cloud

Cloud computing este un concept modern în domeniul computerelor și informaticii, reprezentând un ansamblu distribuit de servicii de calcul, aplicații, acces la informații și stocare de date, fără ca utilizatorul să aibă nevoie să cunoască amplasarea și configurația fizică a sistemelor care furnizează aceste servicii.

Deoarece există o cantitate mare de date generate de diverse aplicații web care sunt dificil de gestionat în bazele de date Cloud, soluția acestei probleme este de a împărți datele. În această lucrare sunt descrise trei tehnici folosite pentru partiționarea datelor pentru aplicațiile de tip Cloud.

The three Vs of Data Storage

1. **Volum** – Cât de mult vom stoca în final? Câtiva GB? O sută de GB?
2. **Viteză** – Care este rata la care vor crește datele noastre? Este o aplicație internă care nu generează un volum mare de date? Este o aplicație externă unde clienții vor încărca imagini și videoclipuri?
3. **Varietate** – Ce fel de date vom stoca?

Este foarte important să ținem cont de acești trei factori în vederea alegerii unui tip de sistem de partiționare care să permită aplicației noastre să se adapteze într-un mod eficient.

Așadar, există în principiu 3 abordări de împărțire:

- partiționarea pe orizontală;
- partiționarea pe verticală;
- partiționarea hibridă;

1. Partiționarea pe orizontală (sharding)

Această tehnică presupune împărțirea unui tabel pe rânduri. De exemplu, dacă avem o bază de date vastă de clienți, o putem împărți în 4 noi tabele: A-G, H-N, O-U, V-Z. Totuși, aceasta nu este cea mai bună soluție, de exemplu, dacă avem de 2 ori mai mulți clienți în gama O-U decât în gama A-G. Atunci când împărțim datele, trebuie să evaluăm numărul de rânduri din noile tabele, astfel încât, fiecare tabel să aibă același număr de clienți, iar în viitor, să crească cu un număr similar de clienți noi.

De asemenea, putem diviza setul de date pe baza clienților recenți, de exemplu, clienții care nu sunt activi în magazinul nostru să fie stocați într-o bază de date. În plus, baza de date care conține clienții activi ar putea fi împărțită în mai multe tabele, pentru a obține rezultate mai rapide.



Structura tabelului original rămâne aceeași în noile tabele, adică avem același număr de coloane.

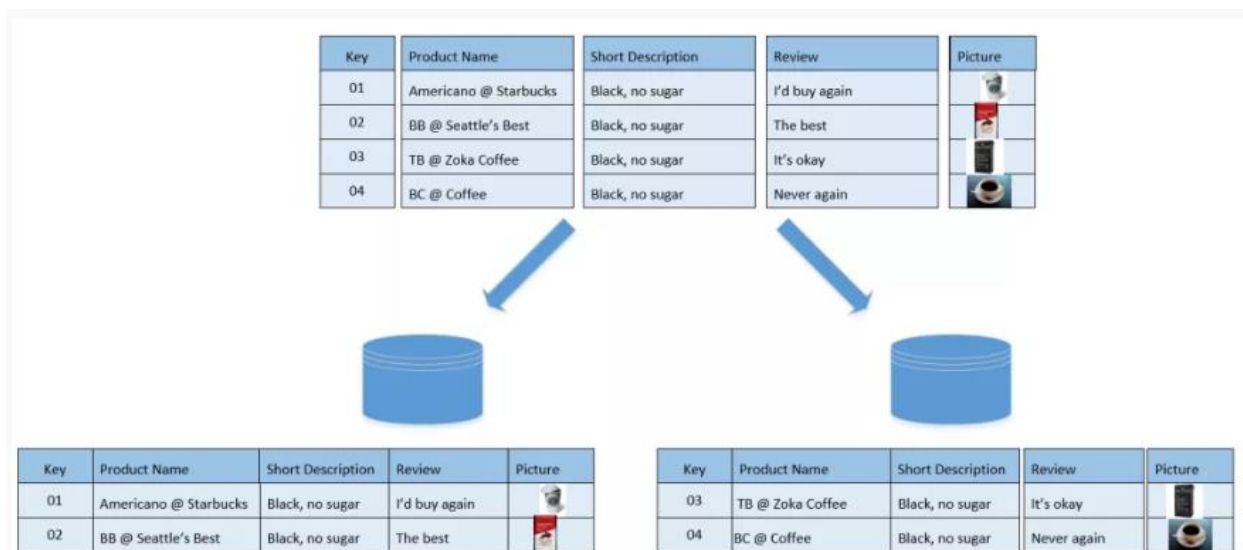


Fig. 1. Partiționarea pe orizontală

2. Partiționarea pe verticală

Acest tip împarte tabelul pe verticală, ceea ce înseamnă că structura tabelului principal se modifică. Un scenariu ideal pentru acest tip de partiție este atunci când nu avem nevoie de toate informațiile despre client în interogarea noastră. De exemplu, dacă avem nevoie doar de comenzile din anul curent, putem să împărțim în două baza de date. Una va păstra informații despre clienți și cumpărăturile recente, iar cealaltă va păstra date despre cumpărăturile din anii precedenți.

Alt exemplu unde partiționarea verticală este o opțiune potrivită este atunci când avem tipuri diferite de date în baza noastră de date (nume sau imagini). Putem păstra valorile în SQL DB și imaginile într-un Azure Blob.

Figura 2 ilustrează reprezentarea vizuală a unui scenariu în care utilizatorul nu are nevoie să vadă scurta descriere a unui produs.

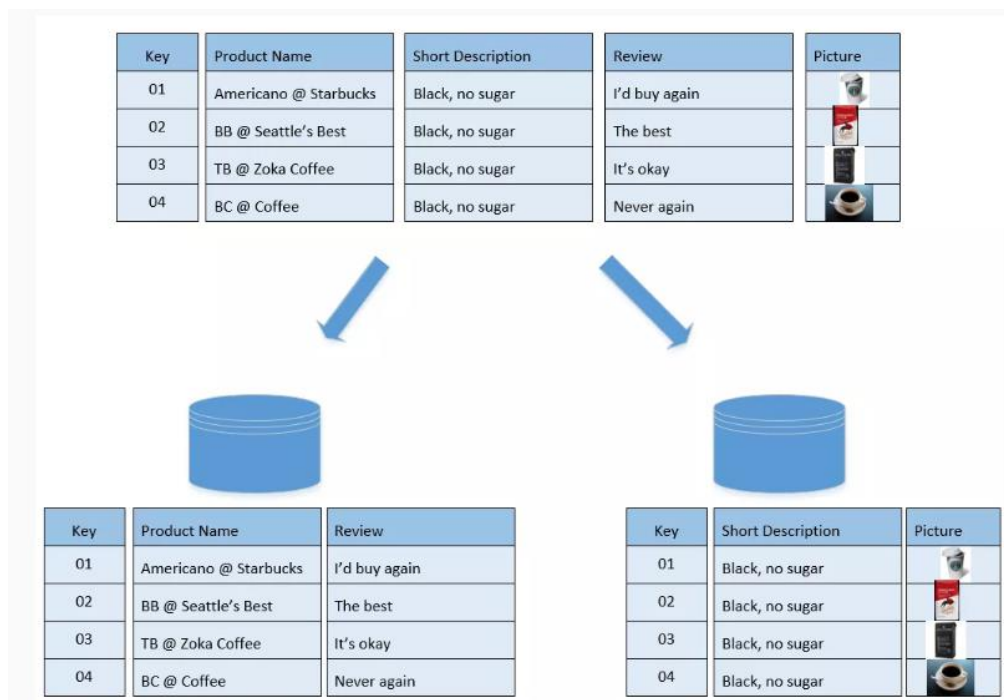


Fig.2. Partiționarea pe verticală

3. Partiționarea hibridă

Această abordare combină cele două tehnici prezentate.

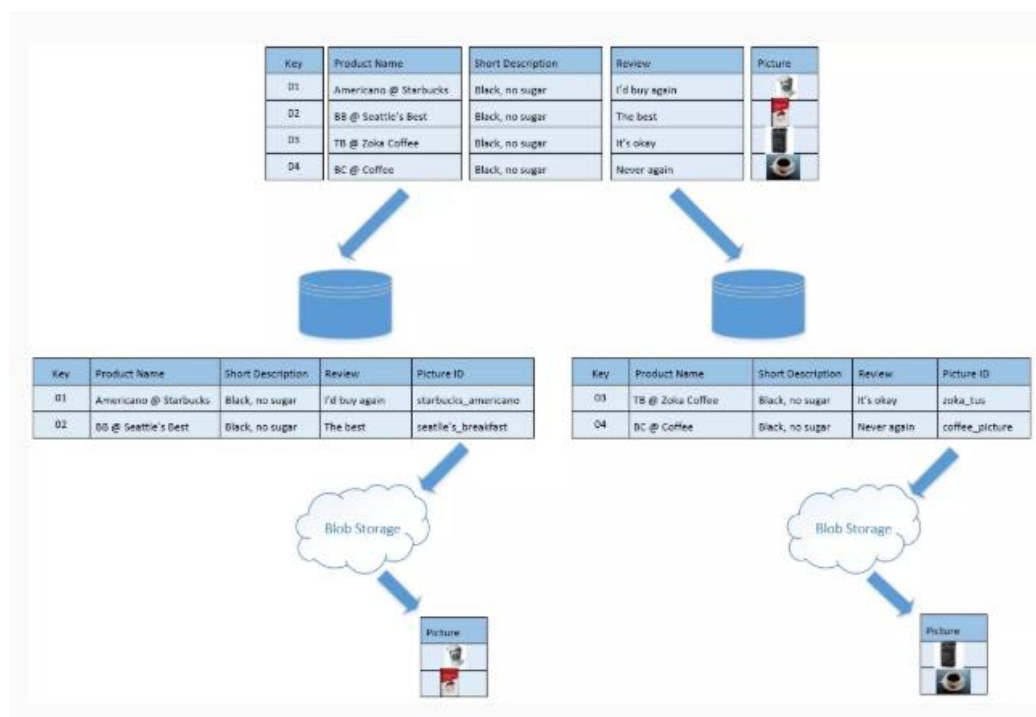


Fig. 3. Partiționarea hibridă



Avantajele oferite de partiționarea datelor:

- Îmbunătățim scalabilitatea
- Îmbunătățim performanța
- Îmbunătățim securitatea
- Asigurăm flexibilitate operațională
- Îmbunătățim disponibilitatea

În concluzie, este necesar să ținem cont de tipul și structura datelor noastre pentru a alege cea mai potrivită abordare prezentată în această lucrare.