

## Tema 2 – DATC

### Partiționarea datelor unei aplicații de tip cloud pe Orizontală

Pentru a înțelege mai bine Partiționarea pe Orizontală vom explica pe scurt la ce se referă în general partiționare.

#### Partiționare – Prezentare și avantaje

În majoritatea aplicațiilor la scară mare datele sunt împărțite în partiții pentru a putea fi gestionate și accesate cu mai mare ușurință. O partiționare corectă ajută o aplicație să se poată ocupa de mai multe date fără să complice accesul la acestea și îmbunătățește performanțele în lucrul cu datele stocate.

În continuare vom expune avantajele partiționării fără a intra în detalii despre acestea: îmbunătățește scalabilitatea, îmbunătățește performanțele, îmbunătățește securitatea, oferă flexibilitate operațională, potrivește locul de stocare a datelor cu modelul de utilizare, îmbunătățește disponibilitatea.

#### De ce să folosim Partiționarea pe Orizontală?

Pentru a înțelege de ce este benefică folosirea partiționării pe orizontală trebuie să vedem limitările pe care le are un server ce stochează informația:

- **Spațiul de Stocare** – pentru o aplicație cloud la scară mare este de așteptat ca volumul de date să fie mare, și să se mărească pe parcursul timpului. Un server de obicei permite extinderea acestuia prin adăugarea de discuri noi sau schimbarea lor cu altele cu o memorie mai mare, dar, până la urmă, sistemul își va atinge limita
- **Resursele de calcul** – o aplicație cloud trebuie să permită un număr mare de clienți să o acceseze în același timp. Puterea de calcul poate fi îmbunătățită prin schimbarea procesorului, dar și acesta are o limită
- **Lățimea de bandă a rețelei** – un singur server, oricât de performant, poate funcționa doar atât de repede cât îi permite transmiterea de date pe rețea, fiind limitat de viteza acesteia. Un trafic mare poate duce la supraîncărcarea rețelei.
- **Locația** – uneori e nevoie ca datele să fie stocate în aceeași regiune cu user-ul din considerente legale, sau pentru a îmbunătăți performanțele

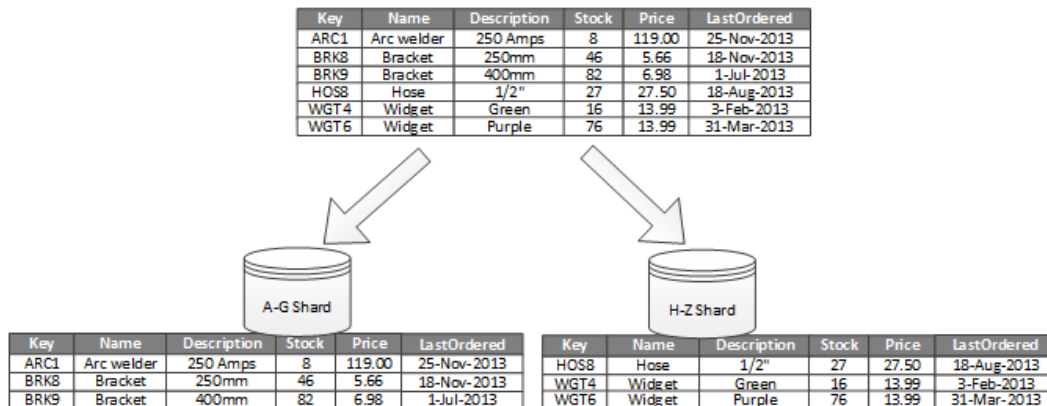
Extinderea prin mărirea capacității de stocare de date, îmbunătățirea puterii de procesare și a conexiunii de internet sunt soluții ce pot au efect temporar.

#### Soluția

Împărțirea datelor în partiții orizontale. Fiecare partiție are aceeași structura (același cap de tabel, aceleași tipuri de date stocate), dar stochează diferite subseturi de date. Partiționarea se face prin seturi de rânduri împărțite din anumite considerente (folosind o *cheie de partiționare*). Coloanele unui tabel rămân neschimbate, iar rândurile sunt împărțite în două sau mai multe tabele, în funcție de necesitate.

Deci, tabelele rezultate prezintă același număr de coloane ca tabela originală, dar pot avea un număr diferit de rânduri una față de alta. Însurarea numărului de rânduri ale tabelelor rezultate în urma împărțirii duce la numărul de rânduri ce îl avea tabelul original.

Spre exemplu datele pentru evidența populației din România pot fi împărțite pe județe (județul fiind *cheia de partiționare*), fiecare județ având aceleași date despre oameni, dar având doar oamenii din județul lui. În următoare figură avem un exemplu în care datele au fost împărțite în ordine alfabetică, datele de la A la G au fost plasate într-o tabelă, iar cele de la H la Z într-o a doua tabelă (*cheia de partiționare* fiind grupul de litere).



Cel mai important factor în împărțirea unei partiții pe orizontală este alegerea *cheii de partiționare*. Este foarte dificilă schimbarea unei *chei* după ce sistemul a fost pus în funcțiune. *Cheia* trebuie să se asigure că distribuie accesările asupra partițiilor cât de egal se poate. Pe lângă aceasta, *cheia* nu ar trebui să fie aleasă pe un set de date ce poate fi schimbat (ex. *vârsta*, *preț*, etc.). Din acest considerent, partițiile nu trebuie să fie egale din punct de vedere al rândurilor. O partiție poate conține multe informații ce sunt accesate rar, iar altă partiție poate conține mult mai puține informații, dar care sunt accesate într-un mod constant și des.

Un alt lucru important este evitarea creării de partiții care să fie foarte des accesate. Spre exemplu, împărțirea după prima literă a numelui provoacă o astfel de partiție, pentru că unele litere sunt mai des folosite decât altele.

Un alt factor pentru a partiționa pe orizontală eficient este împărțirea astfel încât partițiile create să poată fi la rândul lor partiționate ușor, în cazul în care este nevoie ca una, sau mai multe, partiții să fie din nou împărțite.

### Pentru alegerea *cheii* se poate folosi una din următoarele trei strategii

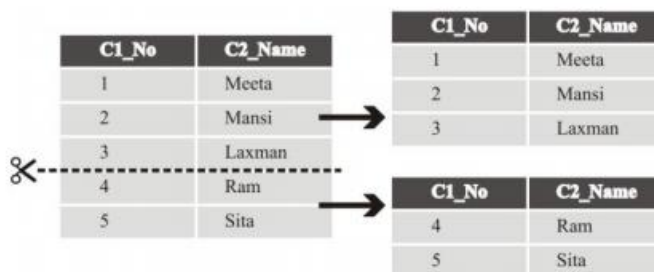
1. **Strategia de căutare** – se implementează o hară ce direcționează o cerere de date către partiția ce conține aceste date, folosindu-se de *cheie*. Spre exemplu, într-o aplicație ce conține mai mulți studenți, datele pentru un student sunt stocate într-o partiție folosindu-se numărul matricol al studentului pe post de *cheie* pentru partiție, deci mai mulți studenți pot folosi aceeași partiție (spre exemplu numerele matricol de la 9000 la 9500), dar datele unui singur student nu vor fi împărțite în mai multe partiții

2. **Strategia de perioadă/interval** – grupează elemente asociate împreună în aceeași partiție iar acestea sunt ordonate după *cheie*; *cheile* sunt secvențiale. Este utilă pentru aplicații în care se interoghează frecvent seturi de date ce se încadrează într-un anumit interval (de exemplu temperatura dintr-o anumită lună – sunt interogate de fapt toate înregistrările făcute în acea lună)
3. **Strategia Hash** – scopul este reducerea încărcării mari pe anumite partiții. Distribuie datele astfel încât să fie păstrat un echilibru între dimensiunea și sarcina de lucru medie de pe partiții. Se creează o funcție ce calculează un hash în funcție de unul sau mai multe atribute ale datelor. Funcția creată trebuie să fie concepută astfel încât să distribuie datele uniform între partiții, eventual și prin introducerea unor elemente aleatorii în calculul hash-ului

### Efectele și avantajele rezultate după partiționarea pe orizontală

#### Efectele Partiționării pe Orizontală

- Tabelele au exact aceeași schemă
- O comandă pe una din tabele este exact la fel pentru toate celelalte tabele
- Fiecare înregistrare reprezintă întreaga informație
- Se pot prelua toate informațiile printr-o singură interogare



#### Beneficiile Partiționări pe Orizontală

- Sistemul poate fi extins prin adăugarea de partiții noi în locații noi de stocare/pe noi servere
- Nu este nevoie de folosirea de echipament specializat pentru stocarea de multe informații, ceea ce reduce și costurile
- Se reduce timpul de conectare și preluare a datelor prin împărțirea volumului de muncă pe mai multe partiții
- În cloud, partițiile pot fi localizate fizic mai aproape de utilizatorii ce accesează datele