

Tema 2 DATC

Partiționarea datelor pentru o aplicație de tip cloud

Cloud computing este mai mult un concept decât tehnologie, tehnologia din spate fiind virtualizarea resurselor și punerea la dispoziție prin închirierea lor. Virtualizarea se face prin folosirea de software precum VMware vSphere. Acesta folosește noi metode de oferire și consumare a serviciilor IT în Internet, servicii care de obicei pot fi dimensionate dinamic și care includ resurse virtualizate. Este de fapt doar o posibilitate secundară, urmare a ușurinței cu care se pot acum accesa toate serverele și centrele de calcul interconectate prin intermediul Internetului. Totodată există o capacitate de scalare atât pe orizontală cât și pe verticală. În cazul dat utilizatorul sau „chiriașul” nu trebuie să mai întrețină echipamente hardware care în anumite condiții pot deveni nefolositoare, toată responsabilitatea este preluată de cel care oferă serviciile date.

Avantaje cloud computing

- *Cost* – oferă un cost redus a resurselor computaționale. Modelul de cloud public are o un model de facturare „pay as you go” prin care plata se face doar pe resursele consumate.
- *Independența de locație* – permite utilizatorului să se conecteze la resurse din orice punct al globului via internet.
- *Mentenanță* – din punct de vedere al clientului nu există nici un efort de mentenanță a infrastructurii, clientului asigurându-se resursele închiriate.
- *Multitenacitate* – resursele fizice sunt distribuite între mai mulți utilizatori acest lucru ducând la creșterea eficienței globale, capacității de scalare, capacității de vîrf, etc.
- *Performanța* – asigură o performanță constantă a sistemului, permițînd construcția unei arhitecturi decuplate prin intermediul serviciilor web
- *Productivitatea* – productivitatea aplicației poate să varieze în funcție de numărul utilizatorilor.
- *Toleranța la defecte* – astăzi furnizorii de cloud computing asigură o valabilitate a resurselor de 99%.
- *Scalabilitate și elasticitate* – unii furnizori de resurse oferă tehnologii de autoscalare a aplicațiilor.
- *Securitate* – cu toate că datele nu sunt ținute „in house”, securitatea poate fi crescută prin centralizarea datelor.

Exista în principiu 3 abordări de împărțire:

1. Partiționarea pe orizontală;

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.

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.

3. Partiționarea hibridă;

Această abordare combină cele două tehnici.

