

Tema 2 – DATC

Document NO-SQL database. Azure Cosmos DB

Nicodin Jesica-Doriana, grupa 4.1, anul 4, I.S

Datorită noilor tipuri de aplicații ce utilizează de regulă o cantitate mare de date complexe, bazele de date relaționale MySQL întâmpină dificultăți, tocmai din cauza lucrului cu cantități mari de date, iar în scopul rezolvării acestei probleme au apărut soluțiile de baze de date cloud ce poartă numele de **NoSQL** (Not only SQL) și sunt baze de date non relaționale.

Ca o definiție a bazelor de date **NoSQL** putem spune că sunt **o nouă generație de baze de date** ce îndeplinesc condițiile: nu sunt relationate, sunt distribuite, open-source, se caracterizează prin scalabilitate orizontală, lipsa unei scheme pentru modelarea bazei de date, suport pentru replicare, API, nu respectă în întregime ACID, stochează o cantitate mare de date. Modul de stocare al datelor în cazul bazelor de date NoSQL nu este prin folosirea tabelelor ci prin folosirea cheilor de identificare, astfel datele pot fi regăsite în funcție de cheile asignate.

Trăsături ale bazelor de date NoSQL:

- Abilitatea scalării pe mai multe servere
- Abilitatea replicării și distribuirii datelor pe mai multe servere
- CLI(call level interface) caracterizat prin simplitate
- Un model concurențial mai slab decât modelul relational(ACID)
- Utilizarea eficientă a indexării distribuite și a RAM pentru o stocare eficientă
- Abilitatea de a adăuga dynamic noi attribute la înregistrările existente

Printre primele aplicații mari ce au întâmpinat problema scalabilității a fost aplicația Facebook, o rețea de socializare cu un număr ridicat de utilizatori, care prin utilizarea bazelor de date NoSQL a avut un beneficiu enorm în urma scalării masive. Sistemul NoSQL implementat de Facebook cu sistemul de fișiere Google se numește “Cassandra”, bazat pe structura BigTable și oferă performanță prin compresie.

Algoritmul matematic ce stă la baza arhitecturii NoSQL este **funcția Hash**. Fiecare cheie fiind partiționată, iar valoarea hash rezultată este folosită pentru a se transmite la un server de baze de date NoSQL. **Scalarea orizontală**, caracteristică a BD NoSQL presupune construirea pe loc și nu vertical și este utilizată atunci când necesitățile aplicațiilor constau în trafic foarte mare și conectarea la un număr mare de evenimente.

Modelele de interogare în NoSQL se bazează pe căutarea unor chei primare sau a unui camp ID și lipsa unei interogări pe alte domenii. (fig.1)

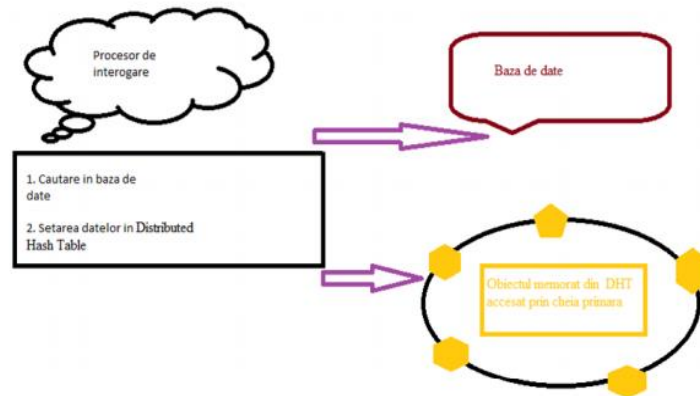


Fig.1

Alte modalități de căutare în NoSQL mai pot fi Scatter/Gather, acestea permit interogarea și indexarea în cadrul unui server de baze de date. Procesorul de interogare poate trimite interogări către nodurile bazei de date, interogările fiind executate local, iar rezultatele de la bazele de date ale tuturor serverelor sunt trimise înapoi la procesorul de interogare. (fig.2)

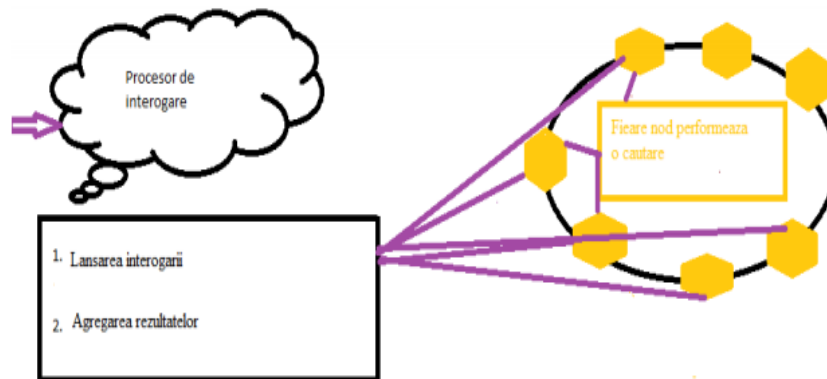


Fig.2

Document NoSQL Database, este un subset al unui tip de bază de date NoSQL, unele magazine de documente pot de asemenea sa fie baze de date cheie-valoare. Acestea sunt utilizate pentru stocarea, preluarea și gestionarea datelor semi-structurate.

Document NoSQL Database, utilizează ca structură pentru stocare și interogări documente. Prin “document” se poate referi la un document Microsoft Word sau PDF, dar de obicei este un bloc de XML sau JSON. În loc de coloane cu nume și tipuri de date care sunt utilizate într-o bază de date relațională, documentul conține o descriere a tipului de date și a valorii descrierii respective. Pentru a adăuga alte tipuri de date într-o bază de date a documentelor, nu este nevoie să se modifice întreaga schemă de bază de date, datele putând fi adăugate prin adăugarea de obiecte în baza de date.

Printre beneficiile utilizării unei baze de date pentru documente se numără:

- Modelarea flexibilă a datelor
- Performanțe rapide de scriere
- Performanțe rapide de interogare

Azure Cosmos DB este un serviciu de baze de date distribuit la nivel global, permite gestionarea datelor chiar dacă acestea sunt păstrate în centre aflate în întreaga lume. Oferă instrumentele necesare măririi modelului de distribuție globală cât și al resurselor de calcul, poate fi folosit pentru modele de documente, valori cheie, relaționale și modele de grafice.

Azure Cosmos DB este mai mult sau mai puțin o bază de date NoSQL deoarece nu se bazează pe scheme, dar pentru că utilizează un limbaj de interogare similar cu SQL și poate susține tranzacțiile ACID a fost clasificat ca și un tip de baze de date NewSQL, însă spre deosebire de alte baze de date NewSQL acesta nu are un model de date relațional.

Oferă un sistem de baze de date la cheie ce poate fi adaptat după propriile nevoi, util celor care caută un sistem de baze de date scalabil și distribuit global(toate resursele fiind împărțite orizontal în fiecare regiune a lumii dar și reproduse în diferite zone geografice).