

TEMA 2 DATC

Document no-sql database. Azure Cosmos DB

Bazele de date tradiționale sunt puse la încercare din ce în ce mai mult de noile tipuri de aplicații care le folosesc. Aceste tipuri de aplicații utilizează de regulă o cantitate mare de date complexe. Dacă în trecut pentru o mare perioadă de timp bazele de date relaționale esențiale pentru aplicații web erau MySQL, astăzi aceste baze de date relaționale întâmpină multe dificultăți în lucru cu cantități mari de date. Pe piața în care activează MySQL au pătruns furnizorii de soluții de baze de date cloud. Aceste baze de date cloud poartă numele de NoSQL- Not only SQL și sunt baze de date non relaționale. Dezvoltarea NoSQL și NewSQL amenință monopolul MySQL.

Datele semistructurate nu sunt organizate conform modelului relational. Ele contin totusi marcaje pentru a separa elementele din care se compun (care se numesc campuri sau attribute - ca in cazul relational). Nu exista o schema a datelor inasa contin in ele descrierea. Pot exista ierarhii de elemente, ordinea campurilor nu este semnificativa. Elemente diferite din aceeasi clasa pot avea attribute diferite.

Bazele de date NoSQL prezintă șase trăsături de bază:

1. Abilitatea de a scala orizontal pe mai multe servere
2. Abilitatea de a replica și distribui datele pe mai multe servere
3. CLI (call level interface) caracterizat prin simplitate (în contrast cu SQL binding)
4. Un model concurențial mai slab decât modelul relațional (ACID)
5. Utilizarea eficientă a indexării distribuite și a RAM pentru o stocare eficientă
6. Abilitatea de a adăuga dinamic noi attribute la înregistrările existente

SQL VS. NOSQL OVERSIMPLIFIED

```
SELECT * FROM Customers_tbl WHERE  
Last_Name='Smith';
```

Cust_No	Last_Name	First_Name
560779	Smith	Juan
207228	Smith	George
173996	Smith	Ben
477610	Smith	Conrad

```
Get customer.firstname,customer.lastname,customer.productID,* where Last_Name='Whitelock'
```

Key	Value
746133	Firstname: George Lastname: Whitelock productID: 2012: 5
135225	Firstname: Luke Lastname: Whitelock productID: 1285: 1 1077: 5
884256	Firstname: Sam Lastname: Whitelock productID: 1442: 2

O altă caracteristică importantă a sistemelor NoSQL este arhitectura „shared nothing” prin care fiecare nod/server este independent, nici unul din ele nu partajează memorie sau spațiu. Datorită acestei caracteristici pot fi efectuate un număr mare de operații de citire/scriere pe secunda.

Azure Cosmos DB este un serviciu de date multi-dimensional, global distribuit, care permite o capacitate mare de stocare a datelor pe regiuni, utilizând latență redusă și disponibilitate ridicată. Avantajele folosirii sunt:

- Distribuție globală: datele sunt distribuite în regiunile dorite cu ajutorul API-urilor, care permit ca datele să fie stocate unde sunt utilizatorii, asigurând o latență scăzută.
- Elasticitate și autonomie la scară de producție și stocare la cerere și la nivel mondial prin mărirea spațiului de stocare și extinderea cu ușurință a capacității bazei de date.
- Dezvoltarea aplicațiilor extrem de receptive.
- Asigurarea disponibilității.
- Dezvoltarea aplicației, distribuită global în mod corect.
- Nu ține cont de o schema a bazei de date sau de gestionarea indexului.
- Costul redus.

Aceste servicii Microsoft folosesc Cosmos DB:

- Office
- Skype
- Active Directory
- Xbox
- MSM

Cosmos DB suportă schemele flexibile și datele ierarhice, fiind potrivit pentru stocarea produselor a căror atribut se pot modifica în timp. Este foarte folosită în industria jocurilor deoarece acestea necesită latențe de o milisecundă pentru a citi și scrie date, Azure Cosmos DB oferă o performanță de elasticitate mărită pentru actualizarea profilului și a statisticilor la zeci de milioane de jucători simultan prin efectuarea unui singur apel API. Indexarea automată permite localizarea jucătorilor prin ID-urile interne ale acestora, fără construirea unei infrastructurii complexe de indexare.