

Nume: Tamas
Horvath An: 4 IS
Grupa: 3.1

Tema 2 DATC: Graph NoSQL database. Azure Cosmos DB

NoSQL ("nu numai SQL"), un tip de baze de date bazate pe grafuri, este o tehnologie pentru gestionarea datelor concepută pentru a gestiona seturi foarte mari de date structurate, semi-structurate sau nestructurate. Acesta tehnologie ajută companiile care se folosesc de foarte multe date să acceseze, să integreze și să analizeze datele respective indiferent de cantitate sau sursa. Bazele de date NoSQL sunt folosite cu predilectie în analiza big data, de exemplu a datelor din cadrul rețelilor de socializare.

NoSQL vs baze de date relationate:

Pana recent, nu existau alte moduri de gestionare a datelor folosite la scara larga in afara de bazele de data relationate ceea ce presupunea stocarea informatiilor intr-un mod structurat(tabele, coloane, randuri). Un astfel de tip de baza de date are implicit nevoie de o schema(modul in care sunt organizate datele si relatiile dintre ele) care sa fie definit inainte ca prima inregistrare sa fie adaugata. Avantajul acestui tip de baza de date este in principiu faptul ca ofera siguranta si stabilitate, deoarece este o tehnologie destul de veche si are limitele cunoscute.

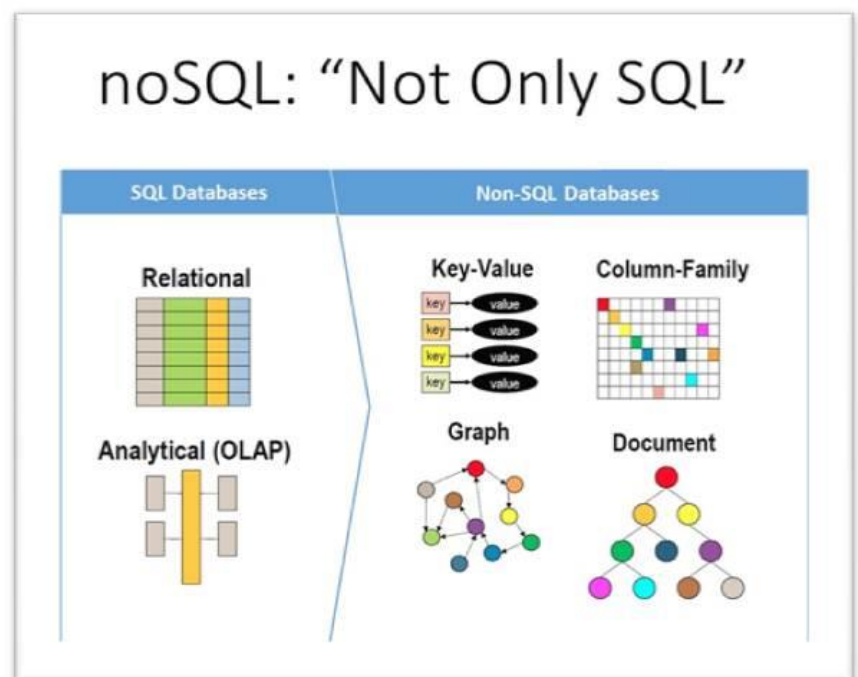
În ultimii ani, datele provenite din device-uri mobile, rețele de socializare și IoT au ajuns să fie întâlnite din ce în ce mai frecvent. Aceste date sunt de cele mai multe ori nestructurate și sunt generate în timp real ceea ce împiedică folosirea abordării tradiționale. Astfel, grafurile NoSQL sunt soluția optimă pentru aceste probleme, ele fiind ideale pentru cantități foarte mari de date și pentru procesarea datelor nestructurate, nefiind nevoie de o schema predefinită pentru adăugarea înregistrărilor. Alte avantaje aduse de abordarea prin grafuri ar fi faptul că baza de date devine mai flexibilă, dinamică și mult mai ieftină din cauza că nu mai este nevoie de o pregătire atât de îndelungată în organizarea datelor.

Cele 2 cele mai relevante caracteristici ale unei baze de date sunt:

- scalabilitatea
- structura

Scalabilitatea:

În principiu, bazele de date SQL sunt scalabile pe verticală, ceea ce înseamnă că modul în care putem să creștem cantitatea de date procesate prin creșterea diferitelor componente(CPU,



RAM, etc). Asta inseamna ca in momentul in care cantitatea de date creste exponential, o baza de date de tip SQL poate sa ajunga depasita. Pe de alta parte, bazele de date NoSQL sunt scalabile orizontal, ceea ce inseamna ca pentru a gestiona mai multe date, adaugam mai multe servere in baza de date. Aceasta metoda permite cresterea mult mai rapida a bazei de date, astfel abordarea fiind preferata in cazul data set-urilor mari si in continua schimbare.

Structura:

Bazele de date de tip SQL sunt bazate pe tabele in timp ce bazele de date NoSQL pot fi stocate in mai multe forme cum ar fi: documente, perechi chei valoare sau grafuri. Din acest punct de vedere, cele de tip SQL sunt mai potrivite pentru aplicatii care au nevoie de tranzactii pe mai multe randuri sau sistemele vechi care sunt construite pentru structuri relationate.

Exemple de baze de date SQL ar fi: MySQL, Oracle, PostgreSQL, and Microsoft SQL Server. Cele mai cunoscute baze de date NoSQL ar fi: MongoDB, BigTable, Redis, RavenDB, Cassandra, HBase, Neo4j and CouchDB.

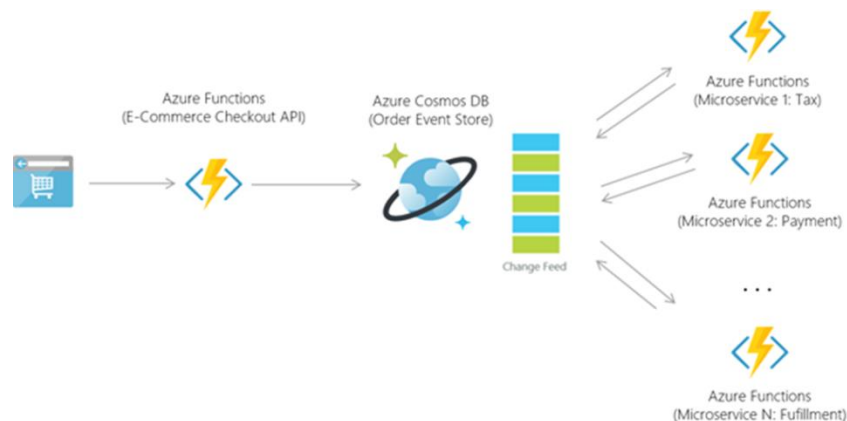
Azure CosmosDB:

Cosmos DB este un serviciu de baze de date care este distribuit global. Permite gestionarea datelor provenite din diferite surse împrastiate în întreaga lume. Poate procesa mai multe modele de date folosind un backend, asta înseamnă că poate fi folosit pentru procesarea mai multor tipuri de modele bazate pe: documente, valori cheie, relaționale și grafice. Deoarece nu se bazează pe niciun fel de schema, această baza de date este considerată NoSQL. Totuși, limbajul de interogare este similar cu cel folosit pentru SQL, este considerat de unii specialiști mai degrabă o baza de date de tip NewSQL.

Problema rezolvată prin această baza de date este nevoia unui mod ușor de a construi și gestiona o baza de date la nivel global, distribuită și scalabilă. În principiu, printr-o abordare clasică, costurile s-ar ridica la o sumă imposibilă de accesat pentru majoritatea companiilor. Prin folosirea Azure CosmosDB, toate acestea pot fi făcute de către o singură persoană cu ușurință, distribuția sistemului ajungând globală. Această distribuție se referă la faptul că toate resursele sunt împartite orizontal în diferite regiuni ale lumii și ulterior reproduse în zonele respective. Astfel, latența bazei de date este minimă ceea ce oferă o experiență optimă utilizatorilor.

Alte funcții de care beneficiază utilizatorii ar fi:

- un produs complet folosit prin Azure care permite reproducerea datelor în nenumărate centre răspândite în întreaga lume;
- serviciu multi API, ceea ce înseamnă datorită indexării automate a datelor, utilizatorii pot accesa datele folosind o multitudine de API-uri. Pot să-și acceseze datele folosind: SQL, Gremlin, JavaScript, Azure Table Storage, and MongoDB;
- folosește mai multe nivele de consistență pentru a asigura integritatea datelor;
- latența foarte scăzută, garantată la mai puțin de 10 milisecunde pentru citire și 15 milisecunde pentru scrierea datelor.



Alte alternative pentru aceasta baza de date ar fi:

- Amazon Relational Database Service (RDS)

- MongoDB Atlas

- Google Cloud Datastore

- IBM Cloudant

- Caspio