Thomas Amorsen
Databases for Developers
10-05/2023

A) What are the advantages and disadvantages of using graph databases and which are the best and worse scenarios for it?

Graf databaser består af "noder" og "relationer" hvor noder er entiteter og relationer beskriver den relation der er mellem noder. I modsætning til f.eks. en relationel database hvor dataen vil være i form af nogle skemaer er en grafdatabase skruet sammen som et netværk. Dette netværk af noder kan derfor indeholde komplekse relationer, der samtidig er forholdsvis billigt at forespørge på. Denne evne til at persistere komplekse relationer gør graf databasen ekseptionel til at analysere brugeradfærd, sociale netværk mm. Det betyder dog ikke at graf databaser er perfekte. Graf databaser understøtter f.eks. ikke transaktioner særligt godt og kræver i mange tilfælde mere cpukraft og lagerplads. Derfor vil en relationel database være bedre i scenarier hvor der ikke er brug for komplekse relationer.

B) How would you code in SQL the Cypher statements you developed for your graphalgorithms-based query, if the same data was stored in a relational database?

Hvis jeg tager **'Match(s:Player)-[r:PLAYS]->(t:Position{name:"CB"})'** der finder alle spillere der spiller positionen 'CB'. Denne data kan blive persisteret på
flere måder i en relationel database men jeg vil gå ud fra følgende. 3 skemaer, en der indeholder spillere, en der indeholder alle positioner og en der
kobler de to skemaer sammen. For at hente samme data som i Cypher eksemplet oven over ville jeg skulle 'join' de 3 tabeller sammen og med et 'where' filtrere i dataen
SQL sætningen vil se nogenlunde således ud:
"'**SELECT * FROM Players join Player_Positions on Players.ID = Player_Positions.PlayerID join Positions on Positions.ID = Player_Positions.ID where Positions.name = 'CB'**'"

C) How does the DBMS you work with organizes the data storage and the execution of the queries?

Neo4j opbevarer dataen i 4 dele:

- Noder
- Relationer
- Properties
- Index

Noder og Relationer refererer dens første property og resten af properties er gemt i en linked liste hvor hver property har en key og en value. Relationer refererer der ud over også deres start og slut node og de noders forrige og næste relation.

DBMS står også for at finde frem til den hurtigste måde at besvare forespørgsler på ud fra de indexes der er tilgængelige.

Thomas Amorsen
Databases for Developers
10-05/2023

D) Which methods for scaling and clustering of databases you are familiar with so far?

Causal clustering er en metode til at håndtere scaling, consistency og safety. Dette cluster opererer med 2 slags servere, en primær og en sekundær.
Den primære håndtere reads og writes, sekundære håndtere kun reads. Det er på denne måde at scaling bliver muligt ved at dyre read operationer kan videresendes til sekundære servere.
For at sikre consistency altså at en bruger kan læse den nyligt indsatte data umiddelbart efter oprettelse af noget er ved at transaktioner har et
'bogmærke'. Dette bogmærke sørger for at en bruger der tidligere har indsat noget data bliver sendt til en server hvor de nyeste writes fra denne bruger er persisteret.