Baze de date și regăsire de informații

Teorie

O bază de date este o colecție organizată, care poate fi citită de mașină, de simboluri, interpretată ca o relatare exactă a unei întreprinderi. O bază de date poate fi actualizată de mașină și, prin urmare, trebuie să fie și o colecție de variabile. O bază de date este de obicei disponibilă unei comunități de utilizatori, cu cerințe posibil diferite.

Design

Proiectarea bazei de date presupune organizarea datelor conform unui model specific de baze de date. Proiectantul are rolul de a determina tipurile de date ce trebuie stocate și modul în care acestea interacționează între ele. Având aceste informații, proiectantul poate începe să structureze datele în cadrul modelului bazei de date. Un sistem de gestionare a bazelor de date se ocupă de administrarea și manipularea datelor în mod corespunzător.

În cadrul procesului de proiectare a bazei de date, se efectuează clasificarea datelor și se identifică interrelațiile dintre ele. Această reprezentare teoretică a datelor este cunoscută sub numele de ontologie. Ontologia reprezintă fundamentul teoretic al designului bazei de date.

Detereminarea stocării datelor

De obicei, persoana responsabilă de proiectarea unei baze de date este un specialist în proiectarea bazei de date, având experiență în acest domeniu, mai degrabă decât cunoștințe specifice despre datele care urmează să fie stocate, cum ar fi informații financiare sau biologice. Prin urmare, identificarea datelor care trebuie stocate în baza de date necesită colaborarea cu un expert în domeniul respectiv, care să fie conștient de informațiile relevante ce trebuie incluse în sistem.

Acest proces este considerat în mod obișnuit parte a analizei cerințelor și implică abilități din partea proiectantului bazei de date pentru a obține informațiile necesare de la specialiștii din domeniu. Acest lucru se datorează faptului că specialiștii în domeniu pot să nu exprime în mod clar cerințele sistemului pentru baza de date, deoarece nu sunt obișnuiți să gândească în termeni de elemente de date discrete ce trebuie stocate. Identificarea datelor care urmează să fie stocate poate fi realizată prin specificația cerințelor.

Dependența/influența de alte subdomenii

Bazele de date sunt strâns legate de alte domenii ale informaticii și există o dependență reciprocă între ele. Iată câteva exemple de cum bazele de date depind de alte domenii ale informaticii:

1. Algoritmi și structuri de date: Algoritmii și structurile de date sunt esențiale pentru operațiile de căutare, sortare și gestionare eficientă a datelor în cadrul bazelor de date. Optimizarea algoritmilor și a structurilor de date utilizate în interogările și manipularea datelor poate duce la performanțe mai bune și la timp de răspuns mai scăzut în cadrul bazelor de date.

1. Limbaje de programare: Bazele de date sunt interacționate și manipulate utilizând limbaje de programare specializate pentru bazele de date, cum ar fi SQL (Structured Query Language) sau limbaje specifice anumitor baze de date. Cunoașterea și utilizarea corectă a acestor limbaje de programare este esențială pentru a opera și dezvolta aplicații care utilizează baze de date.

1. Arhitectura computerului: Bazele de date sunt dependente de arhitectura computerului, deoarece aceasta influențează modul în care datele sunt stocate, accesate și procesate în memoria computerului. Cunoștințele despre arhitectura computerului pot contribui la optimizarea performanței bazelor de date prin alegerea adecvată a hardware-ului și configurarea sistemului de stocare.

1. Sisteme de operare și rețele: Sistemele de operare și rețelele joacă un rol esențial în gestionarea și comunicarea cu bazele de date. Bazele de date rulează pe sisteme de operare și utilizează mecanismele de acces la fișiere și gestionarea memoriei oferite de acestea. De asemenea, bazele de date pot fi distribuite pe mai multe noduri într-o rețea și au nevoie de rețele pentru transmiterea datelor între aceste noduri.

1. Ingineria software: Bazele de date fac parte din multe aplicații software și sunt dependente de procesele și metodologiile de dezvoltare software utilizate în ingineria software. Proiectarea, implementarea și mentenanța bazelor de date necesită o abordare sistematică și aplicarea celor mai bune practici din ingineria software.

1. Inteligența artificială și roboții: Bazele de date sunt utilizate ca sursă de date pentru antrenarea modelelor de inteligență artificială și pentru stocarea rezultatelor acestora. În plus, roboții și sistemele autonome pot utiliza baze de date pentru stocarea și accesul la informații necesare pentru luarea deciziilor și funcționarea lor.

1. Știința computatională: Bazele de date sunt esențiale pentru cercetarea în știința computatională, deoarece oferă un cadru pentru stocarea și manipularea datelor utilizate în experimente și simulări. Analiza datelor și extragerea de informații relevante din bazele de date sunt aspecte cheie ale științei computationale.

1. Grafica: Bazele de date pot stoca și gestiona date grafice, cum ar fi imagini, grafice sau modele 3D. Aceste date pot fi utilizate în aplicații de grafică computerizată, cum ar fi jocuri, simulări sau aplicații de design asistat de calculator. Grafica computerizată poate beneficia de baze de date pentru stocarea și accesarea eficientă a resurselor grafice.

1. Interacțiunea om-computer: Bazele de date sunt folosite în sistemele de interacțiune om-computer pentru a gestiona informațiile relevante pentru utilizatori și pentru a asigura o interfață eficientă și prietenoasă. Informațiile introduse de utilizatori sau obținute în urma interacțiunii pot fi stocate în baze de date și utilizate pentru a personaliza experiența utilizatorilor.

1. Informatica organizațională: Bazele de date sunt fundamentale în cadrul informaticii organizaționale, care se ocupă de gestionarea și utilizarea eficientă a informațiilor într-o organizație. Bazele de date sunt utilizate pentru stocarea datelor organizaționale, gestionarea proceselor de afaceri, luarea de decizii și generarea rapoartelor. Informatica organizațională se bazează pe bazele de date pentru a asigura fluxul adecvat al informațiilor și pentru a sprijini operațiunile organizaționale.

1. Bioinformatica: Bazele de date joacă un rol crucial în bioinformatică, care combină domeniul biologiei cu informatica. Bioinformatica utilizează baze de date pentru stocarea și gestionarea datelor biologice, cum ar fi secvențele de ADN, proteinele și informațiile despre gene. Bazele de date bioinformaționale permit cercetătorilor să acceseze, să analizeze și să interpreteze aceste date în vederea înțelegerii proceselor biologice și a dezvoltării de noi tratamente și terapii.

Problemele importante, problemele deschise (cu exemple care să dea o intuiție despre natura problemelor)

1. Scalabilitate: Una dintre problemele majore în domeniul bazelor de date este scalabilitatea, adică capacitatea de a gestiona volumul mare de date și de a asigura performanță în creștere pe măsură ce dimensiunea bazei de date crește. În prezent, cu creșterea exponențială a datelor, se caută soluții eficiente pentru a manipula, a stoca și a procesa aceste volume mari de date.

1. Securitate și confidențialitate: Bazele de date conțin adesea informații sensibile și critice, cum ar fi datele personale, informațiile financiare sau secretele comerciale. Protejarea acestor date împotriva accesului neautorizat, atacurilor cibernetice și pierderilor de date este o provocare continuă. Asigurarea securității și confidențialității datelor rămâne o preocupare constantă în domeniul bazelor de date.

1. Integrarea datelor: O problemă deschisă în domeniul bazelor de date este integrarea datelor provenite din surse diverse și heterogene. Într-un mediu în care datele sunt stocate în sisteme și formate diferite, este dificil să se asigure interoperabilitatea și coerența datelor în cadrul unei baze de date integrate. Integrarea eficientă a datelor rămâne o provocare importantă în domeniul bazelor de date.

1. Eficiența interogărilor: Cu creșterea dimensiunii și complexității bazelor de date, optimizarea interogărilor devine o provocare. Proiectarea de algoritmi și tehnici eficiente pentru interogarea și recuperarea rapidă a datelor reprezintă o problemă deschisă în domeniul bazelor de date.

1. Gestionarea datelor incerte și incomplete: Bazele de date trebuie să facă față și să gestioneze datele incerte sau incomplete, care pot apărea în diverse situații, cum ar fi datele senzoriale sau datele colectate din surse externe. Cum să se manipuleze, să se stocheze și să se interogheze astfel de date reprezintă o provocare importantă în domeniul bazelor de date.

Persoanele importante

* Edgar F. Codd: Considerat "părintele" bazelor de date relaționale, Edgar F. Codd a formulat modelele conceptuale și teoretice care stau la baza bazelor de date relaționale. Contribuțiile sale au pus bazele principiilor fundamentale ale bazelor de date și ale limbajului SQL.

* Michael Stonebraker: Câștigător al premiului Turing în 2014, Michael Stonebraker este cunoscut pentru contribuțiile sale în dezvoltarea sistemelor de gestiune a bazelor de date (SGBD). Acesta a contribuit la dezvoltarea unor SGBD revoluționare, precum Ingres, PostgreSQL și Vertica, și a adus inovații în domenii precum bazele de date distribuite și bazele de date orientate pe coloane.

* Jennifer Widom: Profesoară de informatică la Universitatea Stanford, Jennifer Widom este o autoritate în domeniul bazelor de date și a adus contribuții semnificative în cercetarea și predarea acestui domeniu. Ea este cunoscută pentru activitatea sa în dezvoltarea modelelor conceptuale, limbajelor de interogare și a tehnologiilor web pentru bazelor de date.

* Jim Gray: Câștigător al premiului Turing în 1998, Jim Gray a fost un pionier în domeniul bazelor de date și a avut contribuții semnificative în dezvoltarea sistemelor de baze de date și a bazelor de date distribuite. Lucrările sale au abordat aspecte precum gestionarea tranzacțiilor, recuperarea de urgență și bazele de date pe scară largă.

* Christos Faloutsos: Profesor de informatică la Universitatea Carnegie Mellon, Christos Faloutsos este cunoscut pentru contribuțiile sale în domeniul analizei bazelor de date, în special în domeniul detectării de anomalii și al modelării rețelelor complexe. A fost distins cu numeroase premii și este unul dintre cei mai citați cercetători în domeniul bazelor de date

Forurile importante (reviste, conferințe)

**Reviste:**

1. ACM Transactions on Database Systems (TODS)
2. IEEE Transactions on Knowledge and Data Engineering (TKDE)
3. Journal of the ACM (JACM)
4. VLDB Journal
5. ACM SIGMOD Record
6. Data Mining and Knowledge Discovery (DMKD)
7. Information Systems (IS)
8. ACM Computing Surveys (CSUR)

**Conferințe:**

1. ACM SIGMOD Conference on Management of Data
2. International Conference on Very Large Data Bases (VLDB)
3. ACM SIGKDD Conference on Knowledge Discovery and Data Mining
4. International Conference on Data Engineering (ICDE)
5. IEEE International Conference on Data Mining (ICDM)
6. ACM Conference on Information and Knowledge Management (CIKM)
7. International Conference on Extending Database Technology (EDBT)
8. International Conference on Database Theory (ICDT)