

1. časť

1. (2b) Objasnite pojmy "entita", "atribút", "doména" a uveďte na ne príklady.

Entita – objekt v reálnom svete s vlastnou existenciou (fyzickou napr. zamestnanec alebo konceptuálnou napr. stretnutie), ktorý je predmetom nášho záujmu

Atribút – charakteristická vlastnosť entity

Doména – množina možných hodnôt pre daný atribút, počet hodnôt v doméne je kardinalita

2. (1b) Vysvetlite pojmy primárny a cudzí kľúč a vysvetlite, na čo slúžia.

Primárny kľúč – jednoznačne-určujúci atribút entity, ktorý je pre entitu jedinečný

Cudzí kľúč – prepája súvisiace entity; hodnoty cudzieho kľúča sú hodnotami primárneho kľúča v inej entite

3. (2b) Vysvetlite pojmy "sekundárny index", "klastrovaný index" a ich využitie.

Sekundárny index – usporiadaný súbor záznamov pozostávajúcich z dvoch polí (indexované pole, odkaz na blok). Sekundárných kľúčov môže byť niekoľko (i pre všetky polia záznamu).

Väčšie zrýchlenie prístupu priamo do súboru než primárne indexy. Sekundárne indexy poskytujú logické usporiadanie záznamov.

klastrovaný index - tvorený nad usporiadaným súborom nekľúčových položiek pevnej dĺžky.

Prípad riedkeho indexovania. V indexe je len jedná položka pre každú hodnotu indexovaného poľa.

4. (1b) Vysvetlite, čo znamená pojem normalizácia dátového modelu a vysvetlite, na čo sa používa.

Normalizácia – formálny postup, ktorý aplikuje súbor pravidiel na spojenie atribútov s entitami

Použitie normalizácie - identifikujeme vzťahy medzi entitami, kombinujeme atribúty do relácií, kombinujeme relácie do databázy, používa sa na vytvorenie relačnej schémy

5. (2b) Vysvetlite pojem "transakcia". Na čo sa používa tzv. "žurnál" pri transakčnom spracovaní a objasnite použitie žurnálu s bezprostrednou a odloženou realizáciou zmien.

Transakcia – skupina zmien (zápisov), ktorá sa vykonáva ako celok; ak pri niektorej zmene nastane chyba vrátia sa aj už všetky vykonané zmeny. Potvrdenie je commit a zrušenie rollback.

Žurnál – ochrana transakcií pred poruchou; uchováva poradie transakcií ako sa vykonávali;

Obnova po páde sa môže zabezpečiť rekonštruovaním operácií zo žurnálu od určitého známeho stavu (checkpoint).

Bezprostredná realizácia zmien – Umožňuje provádět modifikace databáze, když je transakce v aktivním stavu (tzv. nepotvrzené modifikace). V případě výpadku je potřeba u nedokončených transakcí vrátit původní hodnoty a u dokončených znovu zapsat nové hodnoty.

Odloženou realizáciou zmien – Atomičnosti je dosaženo zaznamenávaním modifikácií do žurnálu, ale provedení zápisů je potlačeno, než se transakce dostane do stavu částečného potvrzení.

6. (2b) Vysvetlite, čo je to rozvrh v transakčnom spracovaní a definujte dvojfázovú transakciu.

Rozvrh - usporiadanie operácií transakcie v takom poradí, v akom sa vykonávajú

Dvojfázová transakcia - objekt môže byť v jednom čase zamknutý iba pre jednu transakciu, ak transakcia odomkla aspoň 1 objekt, nesmie už žiaden objekt zamknúť

1.fáza – zamknutie objektov, práca s objektami

2.fáza – postupné odomykanie

7. (2b) Objasnite rozdiely medzi konceptuálnym a logickým návrhom databázy.

Konceptuálny návrh - snažíme opísať predmetnú oblasť pomocou všetkých entít, ktoré sa v nej vyskytujú, a všetkých vzťahov medzi týmito entitami; neriešime implementáciu

Logický návrh – nezávislý od hardwaru/software, výsledkom sú relácie, typy atributov, primárne kľúče, cudzie kľúče, bezpečnostný model

8. (1b) Vysvetlite pretečenú organizáciu záznamov v blokoch a uveďte výhody a nevýhody takejto organizácie.

Pretečená organizácia záznamov v blokoch – na konci záznamu v bloku sa ponecháva pointer na pokračovanie záznamu, ktoré sa už nachádza v inom bloku, teda povolíme záznamom „pretiecť do iného bloku“. Využijeme tým všetok priestor bloku, ale spracovávanie záznamov je zložitejšie; iba pri záznamoch s premenlivou dĺžkou.

Nepretečená - ak nepovolíme záznamom pretiecť do nového bloku, používa sa pri záznamoch pevnej dĺžky a zjednodušuje spracovanie záznamov. Je možná aj pre záznamy s premenlivou dĺžkou.

9. (1b) Uveďte, aké postupy pri svojej činnosti používa optimalizátor dotazov.

1. vygeneruje množiu potenciálnych plánov pre SQL dotaz, založený na dostupných prístupových cestách a hintoch
2. odhadne náklady každého plánu na základe predpokladaných zdrojov, CPU, RAM potrebných pre vykonanie dotazu pomocou daného plánu, plány s vyššími nákladmi sa vykonávajú dlhšie
3. porovná jednotlivé plány a vyberie ten s najnižšími nákladmi, výsledkom je plán popisujúci optimálnu metódu vykonania dotazu

10. (1b) Vysvetlite pojem autentifikácia. Aké druhy autentifikácie poznáte?

Autentifikácia – proces overovania prístupu (heslo, certifikát, biometria)

11. (1b) Vysvetlite organizáciu záznamov v súbore použitím hašovania.

hašovací tabuľka - **štruktúra, ktorá asocjuje kľúče s hodnotami**, kľúč je transformovaný

hašovacou funkciou na adresu záznamu - vytvára pre rovnaký vstup zakaždým rovnaký výstup, vznikajú **kolízie** - pre dva rôzne vstupy vytvorí rovnaký výstup

- **statické h.** - hašovací funkcia je **implementovaná pri vytváraní tabuľky**, pri zmene veľkosti tabuľky sa ťažko mení

- **dynamické h. - dynamická modifikácia hašovacej funkcie**, tak aby odrážal zmeny veľkosti tabuľky
- výhody - **rýchle vyhľadávanie**

Otázky z referátov

12. (1b) Objasnite rozdiely medzi datawarehouse a datamartom a vysvetlite, akú úlohu má dátová pumpa?

Datawarehouse - centrálny systém, ktorý v pravidelných intervaloch zhromažďuje údaje z rôznych interných a externých zdrojov, spracováva ich a vytvára medzi nimi vzťahy a tým pôsobí ako databanka pre systémy riadenia

Datamart – Dátová tržnica, DM, Data mart je špecifický, subjektovo orientovaný sklad dát, navrhnutý na základe špecifických požiadaviek menšej množiny užívateľov

Dátová pumpa – vybrať špecifickú časť dát z prevádzkového systému a túto časť prekopírovať do primárnej databázy datawarehouse zahrňajúc čiastočnú alebo aj značnú zmenu dát a hlavne ich "čistenie"

13. (1b) Čo je to OLAP (Online Analytical Processing) a vymenujte implementačné varianty OLAP?

OLAP – Je spôsob (technológia) uloženia veľkých objemov dát do databázy

14. (1b) Vysvetlite pojem "Business Intelligence".

Business Intelligence – je stratégia práce s informáciami, proces získavania, ukladania, analýzy, správy dát. BI systémy sa niekedy nazývajú aj Systémy na podporu rozhodovania.

15. (1b) Čo rozumieme pod pojmom fuzzy logika a čo je jej prínos ?

Fuzzy logika – cieľom je reprezentovať pravdivostné hodnoty nachádzajúce sa v rozpätí úplná pravda a úplná nepravda; priamo v dotazoch je možné používať vágne pojmy, ktoré sa nachádzajú v reálnych dotazoch formulovaných v prirodzenom jazyku.

16. (1b) Čo je to "cloud computing"?

Cloud computing – možno ho charakterizovať aj ako poskytovanie služieb alebo programov uložených na serveroch na Internete s tým, že používatelia k nim môžu pristupovať napríklad pomocou webového prehliadača alebo klienta danej aplikácie a používať prakticky odkiaľkoľvek

17. (2b) Popíšte nové typy databáz. Popíšte aj ich určenie, spôsob realizácie a súčasne využitie. (HBASE, Hypertable, Cassandra, Voldemort).

Hbase: ide o databázu do ktorej môžete v reálnom čase zapisovať/citať a pristupovať k veľkým dátam...cize daka databaza pre velke tabulky ako pisu biliony riadkov a miliony stlpcov sa tam da nabijat.

Hypertable: open source database inšpirovaná Googlovskou Big Table. Napísaná v C++.

Cassandra: Apache Cassandra open source databázový systém navrhnutý na prácu s veľkým objemom dát naprieč veľkými komoditnými servermi, poskytuje vysokú dostupnosť so žiadnymi zlyhaniami.

Voldemort: distribuovaný data store ktorý je dizajnovaný ako key-value store používaný pre vysokú škálovateľnosť úložiska.

TEORIA

Entita - položka, s ktorou možno manipulovať ako s celkom, združenie viacerých s rovnakými vlastnosťami alebo pravidlami

- vec, objekt v reálnom svete so svojou vlastnou existenciou

Atribút - Každú z entít opisujeme pomocou jej charakteristických vlastností.

Doména - Množina, z ktorej vyberáme údaj určujúci atribút entity. Skupina možných hodnôt pre daný atribút; Množina atomických hodnôt; Počet hodnôt v doméne – kardinalita alebo aj mohutnosť domény (cardinality)

Primárny kľúč- hodnoty, ktoré jednoznačne určujú každý riadok danej tabuľky, nesmie mať hodnotu NULL

Cudzí kľúč- Súvisiace tabuľky sa navzájom prepájajú pomocou tzv. **cudzích kľúčov**. Cudzí kľúč predstavuje položku, resp. skupinu položiek, ktorých hodnoty sú hodnotami primárneho kľúča inej tabuľky.

Primárny index- tvoríme nad usporiadaným súborom kľúčových (unique) záznamov pevnej dĺžky. Sám je usporiadaný, má záznamy pozostávajúce z dvoch polí:

- hodnota poľa, podľa ktorej sú položky usporiadané
- odkaz na *blok*, ktorý obsahuje daný záznam

Súbor môže mať len jeden primárny index.

Klustrovaný index- tvoríme nad usporiadaným súborom neklúčových položiek pevnej dĺžky. Cluster index je prípadom riedkeho indexovania. V indexe je len jedna položka pre každú hodnotu indexovaného poľa

Sekundárny index- usporiadaný súbor záznamov pozostávajúcich z dvoch polí:

- indexované pole (neusporiadáva záznamy v súbore)
- odkaz na blok, kde sa nachádza prvý záznam

Sekundárnych kľúčov môže byť niekoľko (i pre všetky polia záznamu). -väčšie zrýchlenie prístupu priamo do súboru než primárne indexy. Sekundárne indexy poskytujú logické usporiadanie záznamov.

Normalizácia- formálny postup, ktorý aplikuje súbor pravidiel na spojenie atribútov s entitami. –identifikujeme vzťahy medzi entitami, kombinujeme atribúty do relácií, kombinujeme relácie do databázy, používa sa na vytvorenie relačnej schémy

Denormalizácia- Proces denormalizácie sa snaží optimalizovať výkon pri čítaní z databázy pridávaním redundantných dát alebo zlučovaním dát.

1:1 - vzniká medzi dvomi tabuľkami, kde jedna hodnota primárneho kľúča v hlavnej tabuľke odpovedá hodnote poľa v druhej tabuľke (napr. tabuľka objednávky, v ktorej každý záznam má väzbu na tabuľku s vyúčtovaním: každá objednávka odpovedá jednému vyúčtovaniu).

1:N - vzniká medzi dvomi tabuľkami, kde jedna hodnota primárneho kľúča v hlavnej tabuľke odpovedá niekoľkým hodnotám poľa v druhej tabuľke (napr. vzťah medzi tabuľkou so zoznamom kupujúcich a tabuľkou objednávky: každý kupujúci totiž môže urobiť niekoľko objednávok, ale naopak každá objednávka patrí jedinému kupujúcemu).

M:N - vzniká medzi dvomi tabuľkami, kde každý záznam (je jedno z ktorej z týchto tabuliek) môže odpovedať niekoľkým záznamom z druhej tabuľky (napr. vzťah medzi objednávkami a produktami: každá objednávka totiž môže obsahovať niekoľko produktov, alebo produkt môže sa môže nachádzať v niekoľkých objednávkach).

Transakcia- skupina zápisov v jednej connection, ktoré musia „prejsť“ ako celok a pri chybe v niektorej z nich sa neuplatnia ani tie, ktoré už boli v rámci transakcie vykonané. Potvrdenie transakcie sa volá commit, zrušenie rollback

Žurnál- je posloupnost záznamů žurnálu (log record) zaznamenávající všechny modifikace databáze. Jednou z metod ochrany transakcí před poruchou je tvorba tzv. žurnálu transakcí.

Operácie v žurnále transakcií (žurnál=log file)
[start_transaction, I]
[read_item, T, X]
[write_item, T, X, stará hodnota, nová hodnota]
[commit, T]
[abort, T]
[checkpoint]

Tieto operácie popisujú čo sa stalo v jednotlivých transakciách, presne v tom poradí, v ako sa operácie v skutočnosti vykonávali. Obnova po páde sa môže zabezpečiť rekonštruovaním operácií zo žurnálu od určitého známeho stavu (checkpoint). Transakčný žurnál môže byť úplne oddelený od primárnej organizácie bázy dát (môže byť uložený napr. na magnetickej páske a pod.)

Stručně řečeno, žurnály jsou záznamy, které uchovávají informace o průběhu transakcí a slouží výhradně k zotavení databáze po chybě (recovery).

Rozvrh- usporiadanie operácií transakcie v takom poradí, v akom sa vykonávajú

Fragmentácia- proces rozdelenia relácií databázy na množinu fragmentov, ktoré sú z hľadiska umiestnenia navzájom nezávislé a môžu byť uložené na rôznych uzloch siete.

Horizontálna fragmentácia- vykonaná pomocou operácie relačnej algebry Výber (Selection) na globálnej relácii podľa databázovej schémy. Každý z fragmentov má kardinalitu relácie takú, že súčet kardinalít všetkých fragmentov dáva kardinalitu pôvodnej relácie.

Vertikálna fragmentácia- vykonaná pomocou operácie relačnej algebry Projekcia (Projection). Každý z fragmentov má kardinalitu totožnú s kardinalitou pôvodnej relácie. Množina atribútov schémy fragmentu je podmnožinou atribútov schémy pôvodnej relácie, pričom každý fragment obsahuje atribúty tvoriace primárny kľúč pôvodnej relácie. To nám umožní rekonštruovať pôvodnú reláciu pomocou operácie relačnej algebry Spojenie.

Konceptuálny návrh- snažíme opísať predmetnú oblasť pomocou všetkých entít, ktoré sa v nej vyskytujú, a všetkých vzťahov medzi týmito entitami. V žiadnom prípade v tejto fáze neberieme do úvahy neskorší spôsob implementácie a do istej miery ani neskoršie obmedzenia technologického charakteru. Týmto môžeme venovať všetku energiu na pochopenie vlastného problému.

Logický návrh - nezávisly od hardwaru/software, výsledkom sú relácie, typy atribútov, primárne kľúče, cudzie kľúče, bezpečnostný model - na opis sa používa relačná schéma

Fyzická návrh - mapovanie logického na konkrétny DBMS a hardware, výsledkom sú tabuľky, kľúče, indexy, constrainty, procedúry vkladania dát...

Dvojfázový uzamykací protokol-

1. Pred spracovaním akéhokoľvek objektu je nutné vyžiadať zámok na tento objekt.
2. Po uvoľnení zámku nad objektom transakcia už nesmie uzamknúť žiaden objekt.

Ak sú všetky transakcie zabezpečené dvojfázovým uzamykacím protokolom, tak všetky možné rozvrhy sú sérializovateľné.

Autentifikácia- proces overovania postupu(heslo, certifikát, biometria)

Optimalizujte dotaz `SELECT R.a,R.b,S.f,S.h FROM REL1 as R, REL2 as S
WHERE R.d=S.e AND R.a=40 AND S.h>10;`

```
SELECT R.a, R.b, S.f, S.h  
FROM (SELECT a,b,d FROM REL1 WHERE a=40) R  
JOIN (SELECT e,f,h FROM REL2 WHERE h>10) S ON (R.d=S.e)
```

Zo starších skúšok

Dátové modely sú prostriedky konceptuálnej schémy popisujúce štruktúru databázy, vzťahy v databáze, sémantiku databázy a konzistenciu databázy.

- HDBS(hierarchicky) - náročnosť zobrazenia M:N, závislosť od fyzickej úrovne
- SDBS(sietovy) - nemožnosť konštruovať ad hoc otázky, vopred definovaná prístupová cesta
- RDBS(relačný) - slabý semantický obsah, množinovo orientovaný prístup, možnosť realizovať ad hoc otázky
- ERDBS(entitno relacny) - ľahké na pochopenie, veľké množstvo sémantiky
- OODBS(objektovo orientovany) - podpora objektov, dedičnosť
- +rozšírený relačný (objektovo relačný)
- DDBS(distribuvany) - množina logicky zviazaných databáz distribuovaných na jednotlivých uzloch prepojených sieťou, vyššia spoľahlivosť, dostupnosť

Architektúra DDBS

- množina navzájom prepojených databáz, ktoré sú umiestnené na rôznych uzloch tak, že užívateľ s nimi manipuluje akoby boli umiestnené v centralizovanej databáze
- Výhody - efektívnosť spracovania, zvýšená dostupnosť, vyšší výkon, rozšíriteľnosť
- Nevýhody - zložitosť, distribúcia riadenia, bezpečnosť

Popíšte NF

- 0NF** - neexistuje pole, ktoré obsahuje viac ako 1 hodnotu
- 1NF** - všetky atribúty sú nedeliteľné
- 2NF** - relácia je v 1NF a všetky atribúty, ktoré nie sú PK, sú na PK úplne závislé
- 3NF** - relácia je v 2NF a všetky neklúčové atribúty sú navzájom nezávislé

Organizácia dát pomocou hašovania

- hašovacia tabuľka** - štruktúra, ktorá asocjuje kľúče s hodnotami, kľúč je transformovaný hašovacou funkciou na adresu záznamu - vytvára pre rovnaký vstup zakaždým rovnaký výstup, vznikajú kolízie - pre dva rôzne vstupy vytvorí rovnaký výstup
- **statické h.** - hašovacia funkcia je implementovaná pri vytváraní tabuľky, pri zmene veľkosti tabuľky sa ťažko mení
- **dynamické h.** - dynamická modifikácia hašovacej funkcie, tak aby odrážal zmeny veľkosti tabuľky
- **výhody** - rýchle vyhľadávanie

Metódy synchronizácie transakcií

1. zamykanie - objekt sa zamkne a je prístupný len pre jednu transakciu
 - **jednofázové** - nezaručuje serializovateľnosť
 - **dvojfázové** - objekt môže byť v jednom čase zamknutý iba pre jednu transakciu, ak transakcia odomkne aspoň 1 objekt, nesmie už žiaden objekt zamknúť
 - 1.fáza – zamknutie objektov, práca s objektami
 - 2.fáza – postupné odomykanie

2. timestamp - transakcii sa priradí timestamp podľa času, konfliktné situácie sa riešia v poradí danom timestampom

Hustý index - Indexy ukazujú na všetky záznamy.

Riedky index - Indexy ukazujú len na niektoré záznamy. Ostatné hľadáme sekvenčne od najbližšieho predošlého záznamu.

Organizácia dát

- záznamy s pevnou dĺžkou - pre každý záznam presne stanovená jeho dĺžka
- záznamy s premenlivou dĺžkou - záznamy ukladané len podľa ich dĺžky a nemajú túto dĺžku presne stanovenú. Ukladajú sa teda hneď za sebou
- záznamy s premenlivou dĺžkou s oddeľovačmi - podobne ako predchádzajúci typ, sú ale oddelené oddeľovačmi, ktoré určujú kde záznam končí a začína ďalší

Bell-LaPadula model

- model popisujúci množinu prístupových práv pre objekty a subjekty, je vyjadrený bezpečnostnými úrovňami
- klasifikácia - množina 4 prvkov - top secret, secret, confidential, unclassified
- subjekty (užívateľ) má oprávnenie a môže sa do systému prihlásiť na akékoľvek úrovni, ktorá je menšia alebo rovná príslušnému oprávneniu.
- každý objekt má priradenú bezpečnostnú úroveň podľa citlivosti obsahu.

Optimalizátor dotazov

- optimalizátor je vstavaný softvér, ktorý určuje najefektívnejšiu cestu na vykonanie SQL dotazu
- **postup:**
 1. vygeneruje množiu potenciálnych plánov pre SQL dotaz, založený na dostupných prístupových cestách a hintoch
 2. odhadne náklady každého plánu na základe predpokladaných zdrojov, CPU, RAM potrebných pre vykonanie dotazu pomocou daného plánu, plány s vyššími nákladmi sa vykonávajú dlhšie
 3. porovná jednotlivé plány a vyberie ten s najnižšími nákladmi, výsledkom je plán popisujúci optimálnu metódu vykonania dotazu

Entity Type – množina entít, ktoré majú rovnaký atribút

zamestnanec -> množina všetkých zamestnancov -> konkrétna zamestnanec má konkrétne hodnoty jednotlivých atribútov

klúčový atribút – atribút, ktorého hodnota je jedinečná pre jednotlivé entity daného entitného typu; hodnota tohto atribútu môže slúžiť na jedinečné identifikovanie entity

entity, ktoré majú klúčový atribút – **regular (strong) entity types**

slabé entitné typy - nemajú klúčový atribút

redundancia – nadbytočnosť údajov, ktorá vzniká, keď sú v databáze nevyužívané alebo duplicitné údaje.

Systém riadenia databázy (SRDB) je súhrn programov na spracovanie databázovej štruktúry a riadenia prístupu k databáze. Zodpovedajúci anglický termín je **Database-Management System (DBMS)**

O relácii (tabuľke) hovoríme, že je v **prvej normálnej forme**, ak sú všetky jej atribúty atomické, t.j. ďalej nedeliteľné. Inými slovami - každý atribút relácie môže mať iba jednu hodnotu, teda nemôže byť jeho hodnotou ďalšia relácia.

Tabuľka je v **druhej normálnej forme**, ak je v prvej normálnej forme a navyše každý atribút, ktorý nie je primárnym kľúčom je na primárnom kľúči úplne závislý.

O relácii hovoríme, že je v **tretej normálnej forme**, ak je v druhej normálnej forme a zároveň všetky jej atribúty, ktoré netvoria primárny kľúč, sú na sebe nezávislé. Ak by existoval atribút, ktorý je závislý na inom atribúte, musíme ho presunúť do novej entity.

POJMY REFERATY

1-2

Data mining - (Dolovanie z dát) sa dá preložiť ako extrahovanie alebo dolovanie znalostí, informácií z veľkého množstva dát napr. z databázy.

Dátový sklad (angl. data warehouse) je centrálny systém, ktorý v pravidelných intervaloch zhromažďuje údaje z rôznych interných a externých zdrojov, spracováva ich a vytvára medzi nimi vzťahy a tým pôsobí ako databanka pre systémy riadenia.

Data mart - Dátová tržnica, DM, Data mart je špecifický, subjektovo orientovaný sklad dát, navrhnutý na základe špecifických požiadaviek menšej množiny užívateľov.

Fuzzy logika - Cieľom je reprezentovať pravdivostné hodnoty nachádzajúce sa v rozpätí úplná pravda a úplná nepravda. Veľkou výhodou fuzzy logiky je, že na ňu nemajú vplyv nepresnosti a

neurčitosti, keďže vie v databáze vyhľadať aj chybné informácie. To patrí k jej hlavným prednostiam.

Select NAME, AGE, INCOME From EMPLOYEE Where
(AGE = middle) and (INCOME >= high)

To znamená, že priamo v dotazoch je možné používať vágne pojmy, ktoré sa nachádzajú v reálnych dotazoch formulovaných v prirodzenom jazyku.

3-4

grid(mriežka) je model, ktorý umožňuje koordinované zdieľanie zdrojov, ku ktorým možno pristupovať na požiadanie, jednoducho a transparentne bez ohľadu na to, kde sa nachádzajú
Grid computing - je kolekcia počítačových zdrojov z viacerých miest na dosiahnutie spoločného cieľa (viac PC prepojených medzi sebou za účelom poskytnutia vysokého výpočtového výkonu).
(Oracle database 11g)

Multimediálne databázové systémy - Zameriavajú sa na spracúvanie neštrukturovaných multimediálnych dát. Spracúvajú metadáta multimedií a dáta, ktoré je možné získať rozpoznávaním multimedií. (my SQLServer)

(alebo) Systémy, ktoré umožňujú správu, manipuláciu a riadenie informácií (uložených na rôznych médiách), sa nazývajú multimediálne databázy. (MDBS)

5-6

OODBMS (objektovo-orientovaný systém riadenia databázy) (OOSRDB) je systém riadenia databázy, v ktorom sú informácie reprezentované vo forme objektov ako v OOP. OODBMS sa odlišujú od RDMS (RSRDB), ktoré sú reprezentované v tabuľkách. Objektovo-relačná databáza je hybrid obidvoch.

ODMG – 93 - Predstavuje nadmnožinu spoločného jazyka OMG Core navrhnutého pre programovacie jazyky, databázové systémy a komunikáciu vzdialených systémov

Business Intelligence je stratégia práce s informáciami, proces získavania, ukladania, analýzy, správy dát. BI systémy sa niekedy nazývajú aj Systémy na podporu rozhodovania.

OLAP (Online Analytical Processing, (vrátane HOLAP, ROLAP, MOLAP)) Je spôsob (technológia) uloženia veľkých objemov dát do databázy. Výsledkom týchto databáz sú teda analýzy, reporty a správy o dátach uložených. Pri **OLAP** ide o jednorázovo nahrávané dáta, nad ktorými sa vykonávajú zložité dopyty, pri **OLTP** sú dáta priebežne a často modifikované a pridávané a to obvyčajne mnohými

Hypercube (hyperkocka) - je spôsob organizácie dát, ktorý rozširuje dvojrozmerné tabuľkové usporiadanie, tak že každá dátová dimenzia je uložená v jednej osi kocky.

Základné operácie v OLAP analýze (myslim že v skuske staci vymenovat)

Drill-down – umožňuje užívateľovi vo zvolenej inštalácii nastaviť nižšiu (jemnejšiu) zoskupenú úroveň. Jedná sa o navigáciu v hierarchii dimenzií smerom k väčšiemu detailu.

Roll-up – ide o opak predošlej operácie. Vo zvolených inštanciách istej zoskupenej úrovne nastavuje vyššiu (hrubšiu) zoskupenú úroveň (menší detail v hierarchii dimenzií).

Pivoting – umožňuje „otáčať“ dátovú kocku, t.j. meniť uhol pohľadu na dáta na úrovni prezentácie obsahu dátového skladu.

Slicing – dovoľuje vykonávať rezy dátovou kockou, t.j. nájsť pohľad, v ktorom je jedna dimenzia fixovaná v istej inštancii istej zoskupenej úrovne.

Dicing – je obdobou „slicingu“, ktorý umožňuje nastaviť filter pre viac dimenzií.

7-8

Zakl. Požiadavky na bezpečnostne služby v DB:

- dovernosť (prístup len autorizovaným entitám)
- integrita (informácia nebola zmenená žiadnou entitou ,ktora na to nemala pravo)
- dostupnosť (info. dostupná kedykoľvek)
- sledovateľnosť (uchované záznamy o prístupoch)

Útoky na DBS:

- DOS
- SQL injection
- Stolen backup tapes
- Privilege escalation
- targeting unpatched database vulnerabilities

SQLJ – množina jazykových výnimiek, ktorá umožňuje využívanie vnorených statických výrazov v JAVA programoch.

Výhody v porovnaní s JDBC:

- vyššie programovacie rozhranie pre statické SQL
- ľahšie citateľný, menej kódu
- jednoduchšie na správu

Nevýhody SQLJ v porovnaní s JDBC:

- mnohé IDE nemajú podporu SQLJ
- SQLJ vyžaduje „preprocessing“ krok

9-10

Natívne XML databázy sú dátové systémy, ktoré umožňujú uloženie dát vo formáte XML. Na ukladanie a dotazovanie XML dát sa používa logický model XML. **Druhy:** textovo orientované, modelovo orientované.

Výhody oproti relačným DB: XML formát, možnosť jednoduchého validovania dát, výrazovo bohatý jazyk, možnosť zachytiť ľubovoľnú dátovú štruktúru, dokumentačný jazyk

Google AppEngine (GAE) - Umožňuje beh webovej aplikácie na infraštruktúre googlu. Ide o to že si vytvoríte web app hodíte ju na google a nepotrebuje sa starať o nič a beží vám to tam.

cloud computing - možno ho charakterizovať aj ako poskytovanie služieb alebo programov uložených na serveroch na Internete s tým, že používatelia k nim môžu pristupovať napríklad pomocou webového prehliadača alebo klienta danej aplikácie a používať prakticky odkiaľkoľvek. Používatelia neplatia (za predpokladu, že je služba platená) za vlastný softvér, ale za jeho použitie. Ponuka aplikácií sa pohybuje od kancelárskych aplikácií, cez systémy pre distribuované výpočty, až po operačné systémy prevádzkované v prehliadačoch, ako je napríklad eyeOS, Cloud alebo iCloud.

Appscale je to isté čo google appengine a podporujú kompatibilitu s googlovskými app.

Nove typy databaz

Hbase: ide o databázu do ktorej môžete v reálnom čase zapisovať/citať a pristupovať k veľkým dátam...cize daka databaza pre velke tabulky ako pisu biliony riadkov a miliony stlpcov sa tam da nabijat.

Hypertable: open source database inspirovaná Googlovskou Big Table. Napísaná v C++.

Cassandra: Apache Cassandra open source databázový systém navrhnutý na prácu s veľkým objemom dát naprieč veľa komoditnými servermi, poskytuje vysokú dostupnosť so žiadnymi zlyhaniami.

Voldemort: distribuovaný dátový úložisko, ktorý je dizajnovaný ako key-value store používaný pre vysokú škálovateľnosť úložiska.

Shard database architecture : ide o horizontálne rozdelenie kde dizajn databázy resp. jej riadky sú ponechané osamote, namiesto toho aby boli rozdelené do stĺpcov. Každé rozdelenie formuje časť toho celku (shardu). Majú výhodu v tom, že sa znižuje počet riadkov a každá časť teda shard databázy beží na inom serveri a tým pádom sa znížili indexy počet riadkov a urchluje sa čas hľadania v databáze. Cize každá časť databázy môže bežať na inej masine inom serveri ale dokopy tieto časti shardy čiastky vytvoria veľký systém databázy, ktorá je rýchlejšia.