

Které příkazy SQL lze v PL/SQL procedurách a funkcích volat bez omezení a které s omezením?

Bez omezení: INSERT, UPDATE, DELETE

Omezení: SELECT

víceřádkové selecty (jedině pomocí kurzoru – atributy: isOpen, Found, RowCount)

Jak lze v PL/SQL deklarovat datový typ?

explicitně jménem (X NUMBER(7,2))

kopírováním typu jiné proměnné (Y X%TYPE)

kopírováním typu sloupce (E Emp.Name%TYPE), řádku (R EMP%ROWTYPE)

Popiště Rule-Based optimalizaci a jaké má nevýhody.

Odvozuje plán ze syntaxe příkazů a existence indexů.

Cena přístupu k podmnožině řádek v tabulce v klesajícím pořadí:

- Full scan (procházení celé tabulky a ověřování podmínky, vhodné při velkém % vyhovujících)
- Index-range-scan (vyhledávání intervalu v indexu a v nich ověření podmínky)
- Unique-index-scan (vyhledávání jediné možné řádky podle unikátního indexu)
- RowID-scan (vyhledávání na základě známé hodnoty fyzického identifikátoru v db)

Nevýhodné při existenci více neunikátních indexů na jedné tabulce, použitím určitého indexu je možné znemožnit optimalizátoru použití výrazu v dotazu.

Popiště Cost-Based optimalizaci.

Založena na statistikách (počet různých hodnot ve sloupci, počet řádek v tabulce), počítá cenu zdrojů provedení operace. Dokáže rozlišit plány i pro různé hodnoty konstant v dotazu.

Jaké znáte typy indexace, kdy indexace pomůže a kdy ne, kdy indexy vytvářet a kdy ne?

Indexy slouží pro zrychlení přístupu k datům, nemění syntaxi ani sémantiku příkazů pro manipulaci s daty.

Indexy nad sloupci/výrazy (mohou urychlit dotazy)

B-tree indexy (hodnoty v listech, které jsou oboustranně linkované)

- použitelné pokud dotaz omezuje prvních k sloupců indexu

Bitmapové indexy (pro každou hodnotu sloupce/výrazu se vytváří bin. Řetězec obsahující 1 pro řádky s danou hodnotou)

- vhodné pro sloupce s nízkou selektivitou

Oracle vytváří automaticky unikátní (pro PK, kandidátní klíče). Důležité je vytvářet pro FK. Jinak by se měly vytvářet jen v případech, pokud výrazně pomohou často kladeným dotazům (jinak zdržují).

Jaké jsou základní metody data miningu (případně uveďte kdy je optimální je použít)?

Rozhodovací stromy (nesmí být příliš hluboký, jinak nemá statisticky významné hodnoty pro rozhodování v nižších úrovních)

Shlukování (Clustering) – každý dokument je vektor (podle obsahu slov), shluky obsahují podobné dokumenty; užitečné pro klasifikaci vyhledávání dokumentů

Asociační pravidla – pravidla četnosti výskytu např. produktu v objednávkách (stanoven práh s pro vyhledávání $count \geq s$)

Jaká konstrukce se používá pro zpracování víceřádkových SELECT v PL/SQL a jaké jsou její dva typy?

Používá se konstrukce CURSOR – předává programu PL/SQL info z databáze

Explicitní kurzor – nutno deklarovat, otevřít, načíst data a uzavřít

Implicitní kurzor - je deklarován a prováděn přímo v těle programu. V něm povoleny pouze příkazy, které vrací jednotlivé řádky nebo žádné. Jsou obsaženy v SELECT, INSERT, UPDATE, DELETE

Jaké jsou rozdíly boolského a vektorového modelu dokumentografických systémů DIS?

U Booleovského modelu je obsaženo N dokumentů v databázi a k nim M termů, které je popisují (dokument je reprezentován termy). Zpracováním dokumentů na vstupu vzniká posloupnost dvojic <dok_id, term_id> (obecně n-tic)

Nesnadná formulace dotazů, nemožnost ohodnotit vhodnost výstupů, všechny termy v dotazu chápány jako stejně důležité

U vektorového jsou dokumenty reprezentovány vektorem vah termů $\vec{d}_i = \langle w_{i,1}, w_{i,2}, \dots, w_{i,m} \rangle$, kde $w_{i,j}$ je míra důležitosti j-tého termu pro identifikaci i-tého dokumentu. Podobnost mezi vektorem dotazu a vektorem dokumentu dán podobnostní funkcí. Dokumenty na výstupu řazeny sestupně dle podobnosti, lze omezit velikost výstupu (u Bool. ne) a specifikovat minimální nutnou podobnost.

Počet vrácených dokumentů na dotaz uživatele (120) z nich počet opravdu relevantních (90) a celkový počet relevantních v celém systému (300). Spočítat přesnost P a úplnost R.

Přesnost – Pst., že získaný výsledek je skutečně relevantní. $P = Nvr / Nv$

Úplnost – Pst., že je ve výsledku zařazen skutečně relevantní dokument. $R = Nvr / Nr$

Nvr (počet vrácených relevantních dokumentů o nichž si tazatel myslí, že odpovídají dotazu)

Nv (počet vrácených dokumentů o nichž si db myslí, že jsou relevantní a odpovídají dotazu)

Nr (počet všech relevantních dokumentů v databázi)

-> $P = 90 / 120 = 3/4$; $R = 90 / 300 = 3/10$

Co je Embedded SQL a kdy je výhodné ho použít namísto klasického SQL?

Vložený SQL je SQL napsaný v jiném programovacím jazyce, zpravidla Pascal, PL/I, Fortran, C, Cobol Java. Princip vnořeného SQL spočívá v tom, že umožňuje do některého z vyšších programovacích jazyků vkládat příkazy SQL označené standardním prefixem. Ty jsou pak překládány prekompilátorem na volání funkcí knihovny.

Jaké jsou typy JDBC ovladačů (popište)?

- 1) využívá ODBC (přes JDBC-ODBC bridge), obtížně konfigurovatelné
- 2) komunikace s nativním ovladačem
- 3) komunikace s centrálním serverem (Network server) síťovým protokolem
- 4) založen čistě na jazyce JAVA, přímý přístup do databáze

Jaké jsou základní typy datového modelu data warehouseu?

- 1) Konceptuální struktury založené na tabulkách organizovaných do hvězdicovitých schémat
- 2) KS založené na hyperkostkách, které reprezentují data jako multidimenzionální strukturu

Co je to data cartridge (popište k čemu je to dobré, případně uveďte příklad)?

Jedná se o uživatelské rozšíření databáze Oracle (rozšíření serveru – manipulace s daty probíhají na něm). Může definovat nové datové typy, indexy. Hodí se ke zpracování a manipulaci s komplexnějšími daty. Příklady: finanční a právní systémy, multimédia, datové řady, statistické výpočty.

Popište rozdíly mezi R, OO, O-R databázemi, srovnajte výhody a nevýhody.

O-R: vrstvy mezi OO aplikací a SQL databází (myšlení v objektech); pro pohodlí programátorů rychlý, udržitelný vývoj aplikací

OO: objektová identita namísto primárních klíčů; výkon na netradičních datech, ale slabý v databázových rysech, lze využívat odkazů/ukazatelů REF

R: databáze založené na relačním modelu (způsob uložení v logickém smyslu); výkon na tradičních datech a na rozsáhlých dotazech

Popište rozdíly OLTP vs OLAP.

OLTP db – normalizované tabulky, optimalizace pro insert, update

OLAP db – odvoz. tabulky, redundantní data, optimalizace pro dotazy, procesní logika ve schématech

Jaké jsou problémy distribuovaných databází? Nehodící se škrtněte + doplňte vynechaná slova..

- složitost (distribuce ----- "*databáze*", distrib. zpracování dotazu a jeho optimalizace, složité globální transakční zpracování, distribuce ----- "*katalogu*", paralelismus a uvíznutí, případná integrace heterogenních dat do odpovídajících schémat, složité ----- "*zotavování*" z chyb)
- cena (komunikace je navíc)
- bezpečnost
- obtížný přechod (neexistence automatického konverzního prostředku z centralizovaných DB na DDB)
- nespolehlivost ...(tohle se mělo škrtnout - to není problém distribuovaných databází)

Jaké jsou možné načasování triggeru v PL/SQL?

Before, After

Jaké vlastnosti má ideální temporální databáze? Je možné k takové ideální databázi dospět?

Minimální rozšíření existujícího datového modelu, souvislá prezentace chování měnícího se v čase, snadná implementace, vysoký výkon. Dosažení v reálu je prakticky nemožné.

Jaké jsou standardní datové typy pro ukládání objemných dat na serveru (jejich podtypy, max.velikost,..)?

LOB – standardní typy pro ukládání objemných dat na serveru (až 4 GB)

Externí podtyp (pouze pro čtení)

BFILE – samostatný binární soubor vně databáze

Interní podtyp

CLOB – znakový typ v univerzální znakové sadě serveru

NCLOB – znakový typ v národní znakové sadě

BLOB – binární typ

Lze v SQL modifikovat pohledy, případně jaké a s jakým omezením?

Pohled - nástroj jak schovat určitá data, před určitými uživateli.

Většina SQL implementací umožňuje pouze modifikace jednoduchých pohledů definovaných nad 1 tabulkou. Modifikace složitých pohledů nejsou jednoduché, většinou je nelze přeložit.

K čemu slouží nepojmenovaný PL/SQL blok?

Blok použitelný pouze 1x v místě jeho definice. Lze ho uvést do příkazové dávky SQL, definiční části triggeru.

Popište běžné OLAP operace.

Slice: jedna dimenze kostky se stanoví napevno a udělá se řez (čtverec)

Dice: jedná se o slice (řez) 2 a více dimenzemi datové kostky (přímka)

Drill Down/Up: technika při které uživatel přesněji specifikuje (down) nebo zobecňuje (up) u hierarchií (např. kraj -> město -> ulice)

Roll-up: počítá různé součty dat příbuzných pro jednu nebo více pevně stanovených dimenzí

Pivot: změna orientace kostky (tzn. změna pohledu na data)

Jaké typy tabulek obsahuje hvězdové schéma a jaký je jejich účel (co je obsahem)

Tabulky faktů, tabulky dimenzí a měřítka.

V těchto tabulkách jsou uložena data, nebo nějaké skutečnosti.

Jaké prostředky má ODL k vyjádření dědičnosti?

Klíčové slovo UNDER vyjadřuje dědičnost (create type t1 UNDER t)

U funkcí a procedur se pak vyskytují modifikátory INSTANTIABLE (fce nemá definovanou implementaci), OVERRIDING (fce redefinuje metodu předka), FINAL (nelze redefinovat v potomcích).

JDBC princip.

Je to rozhraní pro unifikovaný přístup k datům, inspirováno rozhraním ODBC. K připojení k databázi se využívá ovladače (jsou k dispozici pro většinu db systémů), který je zpravidla implementován výrobcem databáze.

Co je to ODBC?

Open Database Connectivity. Specifikace API pro databáze, nezávislé na db a jazyce + databázově závislé ovladače. Ovladače přístupné přes správce - Driver Manager.

Příklad: vytvořit abstraktní datový typ Stipendium

```
Create TYPE Stipendium AS OBJECT (
```

```
...
```

```
);
```

Napsat možnosti, jak lze v PL/SQL dělat cykly.

```
LOOP ...tělo... END LOOP;
```

```
WHILE podmínka LOOP...tělo...END LOOP;
```

```
FOR i IN 1..10 LOOP ...tělo...END LOOP;
```

Co je to aktivní databáze?

Co jsou to semistrukturální data? Uveďte příklad.

Jaké výhody přináší jazyk SQL3 (SQL99) ?

Jaké typy temporálních dotazovacích jazyků znáte ?

Temporální databáze doplňte údaje: modely času (lineární, větvený, _____) modely hustoty (_____, hustý, spojitý).

Výhody distribuovaných databází, napište vlastnosti:

- _____ - uživatel nevidí na kolik počítačů je rozděleno
- _____ - větší výkonnost serveru
- _____ - výpadek počítače neohrozí databázi

Jak zlepšuje propustnost komunikačního kanálu PL/SQL ?

Nějaké příklady (kód) na přidělování práv (GRANT) a napsat kdo má jaká práva.

Doplňte operace (u hyperkostky), ke každé byl krátký text
_____, Drill-down, Roll-up, Pivoting, Operand : _____, Slice_and_dice

Na co musíme dávat pozor při návrhu databáze?