

OLAP On-line analytical processing

OLTP rendszer-Adattárház-Adatpiac-OLAP alkalmazás alapfogalmak (ppt)

Az OLAP a vezetés szemszögéből: Demo a Cube Browser használatával

1. Sales kocka megnyitása
2. függőleges dimenzió: customers, mérték: unit sales
3. lefűrés 1997 hónapjaira
4. lefűrés: észrevevessük, hogy USA, OR, Lebanon, Darla Wong : fontos vásárlónk
5. kérdés, mit és hol vásárol Carla? Customers dimenzó lecserélése a Product dimenzóra
6. product alá behozzuk a store dimenziót
7. a két dimenziót egyszerre fűrjük le: kiderül, hogy Darla Salemben vásárol, és szereti a züldséget, sokkal inkább, mint a bort, és a jó minőséget

Áttekintés

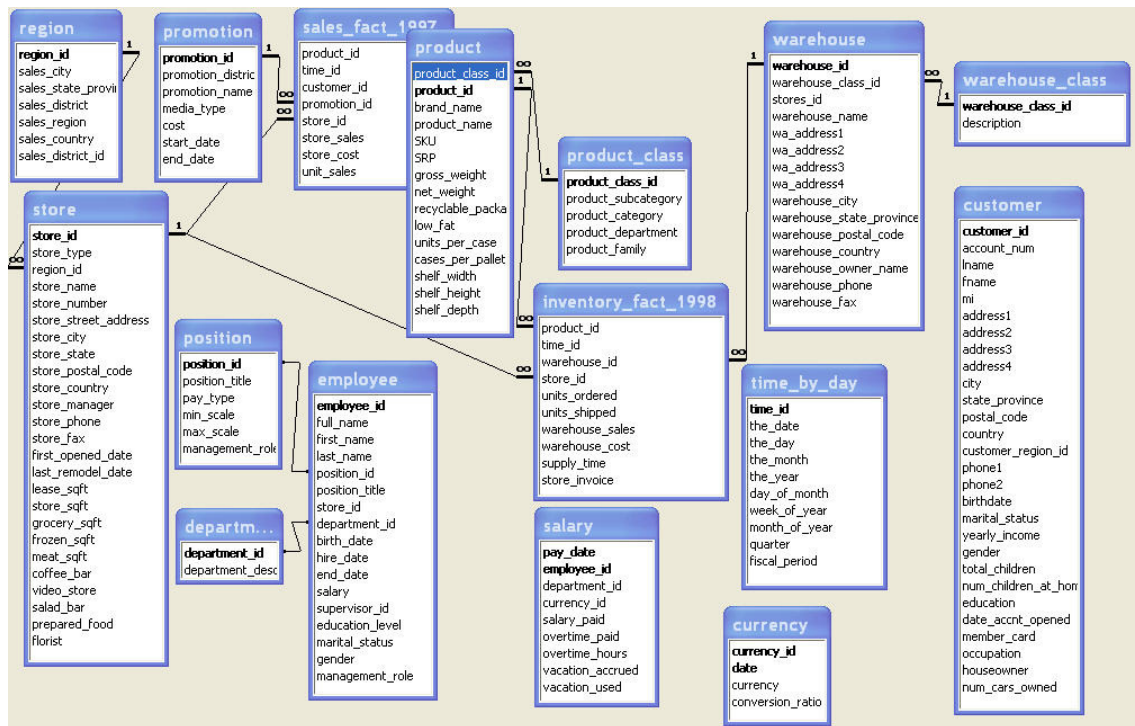
- Elemzési célra leválasztott és újrastrukturált adatok sok OLTP és egyéb forrásból
- Cél az 'ad hoc' lekérdezések gyors kiszolgálása <-> megtervezett, alkalmazásba épített, optimalizált lekérdezések
- Nincs adatmódosítás <-> tranzakciók követése
- Archív, nagy időablakot átfogó adatok <-> az aktuális adatok, szűk időhorizonttal
- A teljesítmény érdekében denormalizálás <-> 3NF
- Nagy méret <-> korlátozott méret a gyors működés érdekében
- A cél a döntéstámogatás (felhasználók a döntéshozók) <-> az üzleti folyamat kiszolgálása
- Az OLTP és az adattárház között: extraction-transformation-loading (ETL) műveletek

- Staging area (az adattárház adatainak is egységesnek kell lennie)
- Tervezési szempontok: adatforrások, transzformációk, frissítési időköz
- Sémák: csillag, hópehely, 3NF, hibrid
 - Csillag: egy nagy ténytábla, több kicsi (denormalizált) dimenziótábla
 - Hópehely: egy dimenzió több táblában
- Ténytábla mezői:
 - Külső kulcsok
 - Numerikus tények (azonos aggregáltsági fokon)
- A dimenziók hierarchiát (szinteket) tartalmaznak, akár egy denormalizált táblán belül
- A szűk keresztmetszet az I/O, nem a tárolókapacitás

MS Analysis Server (2000) architektúra áttekintése

OLAP-tutorial MS Analysis serverrel: foodmart

- Sales_fact_1998 ténytábla, store_cost, unit_sales, store_sales mértékek
- Time_by_day tábla the_date dimenzió,
- Product, product_class tábla, product_category, product_subcategory, és brand_name mezők
- Customer tábla, Country, State_Province, City, és lname mezők



Önálló munka 1: Northwind adatbázis Order Details, Orders, Employees, Products, Categories táblái alapján egy adatkocka készítése

- Dimenziók: product (cat.name, product name), order date (idő típusú), employee (beosztás, név)
- Kocka, adatforrás létrehozás, dimenziók, böngészés

Önálló munka 2: A feladat egy adattárház és egy OLAP-kocka megtervezése és létrehozása a pubs adatbázis alapján. A kocka **mértéke** az eladott könyvek összes és maximális mennyisége. **Dimenziók:**

- a könyv/kiadó neve (szintek: pub_name, type, title)
- eladás helye (szintek: stor_name, state, city)
- kiadás helye (szintek: pub_name, city, country, state)
- eladás ideje (ord_date, szintek: év, negyedév, hónap, nap)
- kiadás ideje (pub_date, szintek: év, negyedév, hónap, nap)
- (akinek megy): a könyv árkategóriája a price mező alapján (10000 és 20000 között „közepes”, alatta „olcsó”, felette „drága”). Segítség: SELECT CASE...

1. feladat: a szükséges táblák: stores, sales, titles, publishers importálása a saját adatbázisba (kihagyható sa jogokkal).
2. feladat: adattárház készítése denormalizálással a fenti táblákból: a titles, és a publishers táblákból hozzunk létre egy olyan táblát mely a számunkra fontos mezőket tartalmazza. Segítség: a title_id is fontos lesz, mert ez köti a dimenziótáblát a ténytáblához. Használjuk a select ... into... szerkezetet, és hozzuk létre kézzel a külső kulcs hivatkozást a title_id alapján.
3. feladat: készítsük el a kockát az Analysis Manager segítségével, és keressük meg, hogy a New Moon Books kiadó legnagyobb forgalmú könyve hol és mikor (melyik városban, melyik hónapban) fogyott a leginkább!

(A helyes válasz ellenőrzésképpen: Tustinban, CA államban 1994 szeptemberében 75-öt adtak el az Is anger the Enemy? c. könyvből, a pszichológia témakörben, konkrétan a Barnums könyvüzletben. Ennek kellene látszani a Cube Browser-ben.)

Ellenőrző kérdés:

Milyen sémájú az elkészült adatkocka? Mi volt a denormalizálás célja?

incremental update vs. full process vs refresh data: demonstrálás új adat beszúrásával. A fenti pubs kocka: qty értékét módosítjuk

- új adat, de nincs új kulcs: incr. update + a hozzáadandó rekordok körét WHERE feltétellel korlátozni kell az új rekordokra, különben minden 2x fog szerepelni!
- új adat, új kulcs (pl. új dátum):: incr. update + incr. update shared dims (különben member key not found hibaüzenetet kapunk)
- Ha újra ki akarjuk számoltatni az összes aggregációt, pl. korábbi adat változás miatt (bár nem szokás adattárházban): refresh data

- Ha újra ki akarjuk számoltatni az összes aggregációt, és új kulcsok is vannak (pl. új dátum): refresh + incr. update shared dims (különben member key not found hibaüzenetet kapunk):

```
--5-tel megnöveljük a mennyiséget új dim. mber hozzáadása nélkül
--az eddigi 15 helyett 20 kell legyen a mennyiség
insert sales (stor_id, ord_num, ord_date, qty, title_id, payterms)
values (6380, 106872, '1999-09-14', 5, 'BU1032', 'xxx')
--stor_name: Eric the Read Books
--title: The 's Guide
--pub_name: Algodata Infosystems
--szűrés nélkül: 15+20=35 (hibás mennyiség)
```

```
--új dim. member (kulcs): új dátum
insert sales (stor_id, ord_num, ord_date, qty, title_id, payterms)
values (6380, 106872, '1999-09-14', 5, 'BU1032', 'xxx')
--megjelent az 1999, már 25 van belőle
```

- új dimenzió: full process

További OLAP

- **Rekurzív szerkezetek** ábrázolása egy dimenzióban, pl. ki kinek a főnöke. A szintek száma az adatoktól függ. Ez egy **parent-child**, szülő-gyerek típusú dimenzió.
- Foodmart-ban egy új dimenziót készítünk: employee tábla, a dimenzió mezői: employee_id, supervisor_id, full_name. Ezekkel egy új, szülő-gyerek típusú dimenziót készítünk emp néven
- Emp dimenzió böngészése (Dimension browser: lehet látni, milyen mély a hierarchia)
- Új kocka, ténytábla: salary, mértéke a salary_paid, vacation_used, number_of_employees = distinct count employee_id
- Dimenziók: emp, store (store tábla store_country, store_state, store_city, store_name szintekkel), time (time_by_day tábla)
- A join-okat kézzel kell betenni, mert az adattárházban nincsenek ilyen külső kulcsok. Konkrétan: employee.store_id -> store.store_id és salary.pay_date -> time_by_day.the_date

- Dimension editor: members with data -> non-leaf data visible (megengedi az adatok delegálását az egyes szintek között)
Különben processzálskor hiba: „A member with key '163' was found in the fact table but was not found in the set of leaf members of dimension 'emp'.”
- Ezután processzálni, de storage options nélkül

Önálló feladat:

Ugyanez a Northwind Employees táblájával (a ReportsTo mező a főnökre mutat)

- **Drill through:** az eredeti OLTP adatokhoz való hozzáférés a kockából (cube editorban engedélyezni tools -> drill through options)
- **Action** (cube editorban definiálható) pl. egy weblap megnyitása az emp dimenzió aktuális eleméhez: insert -> action, target: „a dimension in this cube”. Az url legyen: "http://www.google.hu/search?q=" + [emp].currentmember.name
Ezután az alkalmazott neve felett a jobb gombos menüből indítható az url.

Egyebek:

- Lusta embereknek, akik nem akarnak az adattárházba egy view-t készíteni egy hiányzó mérték számára: Számított tag a sales kockában: átlagos ár = measures.[store sales]/measures.[unit sales] (calculated member builder)
- Member properties: egy dimenzió egy member-éhez csatolt plusz információ (mintha önálló dimenzió lenne. Példa: customer dimenzió (customer tábla) Lname szintjéhez a gender, mint member property. Külön dimenzióként jelenik meg a kockában). Ezekre alapozva lehet **virtuális dimenziót** építeni. (Virtual Dimension: The member properties of another dimension). A virtuális dimenziót hozzá lehet adni egy már létező kockához.

MIS2 ipari alkalmazás architektúrája

- Folyamatvezérlő szerver perces jelekkel
- Adatbázis-terv:
 - OLTP és adattárház egyben, kockákat kiszolgáló táblák és nézetek, táblafrissítés triggerekkel
 - Hópehely struktúra
 - Éves, havi, napi, óra, perc szintű horizontok (táblák) és kockák biztosítása a feldolgozási idők növekedése miatt
- Kockák
 - Külön kockák az extenzív és intenzív típusú jeleknek
 - Számított mérték (átlag)
- Felhasználói felület: MIS2Admin
 - Jelek, horizontok definiálása
- Karbantartás: backup
 - Adatbázis időszakos backup-ok, folyamatos törlés jelhorizontokkal, adott dátumról visszatöltés (aktív/archív adatbázis koncepció)
 - kliens felület a MIS2Admin-ban, szerver oldali támogatás (tárolt eljárások)

- sp_archive_interval: backup job átparaméterezése, paraméter: ennyi naponként készít a mis2_archiv nevű job egy full backup-ot az sp_ment_mai_datumra eljárás meghívásával
 - sp_archive_param: archiválás eredményének lekérdezése
 - sp_szures: napi törlések a horizontok mentén
 - sp_mentesek
- Karbantartás: frissítés
 - Kockák automatikus frissítése
 - DTS csomagok
 - Ezek parancssoros futtatása időzített jobok által
 - kliens felület a MIS2Admin-ban, szerver oldali támogatás:
 - sp_kockafrissit, a job elindítása az adatbázisból;
 - sp_frissitesdatuma, job lekérdezése
 - Naplózás (Kockanaplo tábla)