

Proiectarea depozitelor de date

1. Consultați diagrama conceptuală generată în laboratorul anterior și precizați ce fel de tip de modelare a fost ales pentru implementarea acestui *data mart*:
 - a. modelare entitate-relație;
 - b. modelare multi-dimensională.
2. Identificați tabela de fapte și răspundeți la următoarele întrebări:
 - a. Care este cheia primară?
 - b. Care sunt cheile externe?
 - c. Ce coloane identifică în mod unic o înregistrare?
 - d. Care sunt măsurile (metricile)?
 - e. Ce tip de date sunt utilizate pentru măsuri (date aditive, date semi-aditive, date non-aditive)?
 - f. Ar fi util să se introducă o măsură nouă care să reprezinte valoarea vânzării la nivel de înregistrare? În această ipoteză, ar trebui să se păstreze măsurile care determină această valoare?
3. Identificați tabelele dimensiune și răspundeți la următoarele întrebări:
 - a. Există tabele dimensiune pentru care toate datele sunt generate?
 - b. Există tabele dimensiune pentru care sursa din baza de date *OLTP* este identică?
 - c. Există tabele dimensiune pentru care s-a aplicat un proces de denormalizare?
 - d. Există tabele dimensiune care conțin ierarhii?
4. Precizați care este tipul de schemă ales pentru modelare:
 - a. schemă fulg;
 - b. schema stea;
 - c. schemă multi-stea;
 - d. constelație.
5. Verificați ce înregistrări conține tabela de fapte și răspundeți la următoarele întrebări:
 - a. Ce coloane au valori generate?
 - b. Ce coloane au valori importate din baza de date sursa *OLTP*?

- c. Ce măsuri ar putea fi eliminate, știind că scopul acestui *data mart* este determinarea valorii totale a vânzărilor realizate în funcție de anumiți indicatori?
- d. Puteți da un exemplu de *data mart* pentru care ar fi utile măsurile identificate la punctul c? Ce tabele dimensiune ar trebui păstrate în acest caz?
- e. În ipoteza de la punctul d, considerați că este de preferat să se implementeze un alt *data mart* sau o constelație?
6. Analizați și dați exemple de rapoarte ce se pot obține în următoarele cazuri, identificând modul corect de implementare:
- a. tabela dimensiune *produse* conține atributul *pret_unitar*, iar tabela de fapte *vanzari* nu conține acest atribut;
 - b. ambele tabele (atât tabela dimensiune *produse*, cât și tabela de fapte *vanzari*) conțin atributul *pret_unitar*.
7. Pentru fiecare tabelă dimensiune precizați lista atributelor care:
- a. nu ar trebui actualizate în mod uzual;
 - b. ar putea fi actualizate în anumite situații (dați exemple concrete);
 - c. sunt actualizate frecvent.
8. Analizați toate tabelele dimensiune ale modelului, pentru fiecare dintre acestea precizând dacă:
- a. importurile masive de date le vizează sau nu;
 - b. este de așteptat să se importe sau nu date noi, dar reduse cantitativ;
 - c. este de așteptat să se importe date noi, cu frecvență ridicată sau redusă;
 - d. există tabele dimensiune care pot deveni în timp mai mari decât tabela de fapte.
9. Studiați diagrama conceptuală a bazei de date *OLTP*. Pentru fiecare tabelă din schema *data mart*-ului precizați:
- a. lista tabelelor care conțin datele sursă;
 - b. lista atributelor preluate direct din aceste tabele sursă;
 - c. lista atributelor care constituie bază pentru anumite valori aggregate din tabelele *data mart*-ului;
 - d. modalitatea de obținere a informațiilor necesare populării cu date a tabelelor *data mart*-ului (operații de tip proiecție și/sau selecție aplicate direct pe o tabelă sursă, operații

de tip *union/intersect/minus*, operații de tip *join* aplicabile pe mai multe tabele sursă sau pe aceeași tabelă sursă).

10. În tabela de fapte copie *vanzari_**** adăugați coloana *valoare* care va fi de tip *Number* și nu va permite valori *null*, având valoarea implicită *0*. Actualizați această coloană știind că reprezintă valoarea vânzării unui produs la nivel de factură.
11. Verificați ce date calendaristice conține tabela *tmp*. În ce interval sunt cuprinse acestea?
 - a. Definiți tabela dimensiune copie *tmp_extins_****, având aceeași structură și date ca tabela sursă.
 - b. Se presupune că sunt programate noi importuri de date, ce conțin informații relative la vânzările realizate în perioada 1 ianuarie 2008 – 31 decembrie 2008. Adăugați înregistrări noi în tabela *tmp_extins_**** corespunzătoare acestui interval. Permanentizați tranzacția.
 - c. Ce se întâmplă în cazul în care se constată că în urma importului pe data de 5 iulie 2008 nu există vânzări efectuate? Înregistrarea corespunzătoare acestei date ar trebui eliminată din tabelă?
12. Se presupune că un client își modifică numele. Pentru exemplificare utilizați tabela dimensiune copie *clienti_****, respectiv tabela de fapte copie *vanzari_****. Modificarea vizează clientul având codul 94, numele acestuia devenind *XYZ*.
 - a. Afîsați toate informațiile dimensionale corespunzătoare acestui client.
 - b. Afîsați valoarea totală a vânzărilor pentru acest client, incluzând în rezultat informații descriptive pentru fiecare dimensiune a faptelor.
 - c. Presupunem că în urma importului de astăzi se aplică 2 operații: modificarea numelui clientului la noua valoare și adăugarea de linii în tabela de fapte, corespunzătoare unei singure facturi (*Obs.* Data emiterii facturii trebuie să fie în intervalul calendaristic stocat curent în dimensiunea *tmp*). Aplicați aceste operații direct la nivelul *data mart*-ului. Repetați punctele a și b. Ce observați? Rapoartele generate anterior zilei de astăzi ce valori utilizau? Dar cele de astăzi?
 - d. Dați o soluție de implementare în cazul în care se dorește ca:
 - rapoartele anterioare modificării să utilizeze valorile corespunzătoare clientului la acea dată, iar cele prezente valorile noi, dar să se pună în evidență că este vorba despre același client;

- rapoartele anterioare modificării să utilizeze noile valori corespunzătoare clientului la data curentă, dar să se cunoască evoluția în timp a modificărilor.

13. Folosind ca sursă tabelele *data mart*-ului definit în schema în care v-ați conectat, definiți alte 2 *data mart*-uri a căror model multi-dimesional să fie de tip:

- a. schemă fulg
- b. constelație

Observație: Fiecare nou *data mart* va fi creat în schema altui utilizator, conform informațiilor precizate la laborator)