**METODA DE MODELARE MULTIDIMENSIONALĂ**

**Scopul lectiei**

După ce veti studia materialul acestei prelegeri,

***veți ști:***

* + - * elemente principale ale *modelului multidimensional: atribute*,   *dimensiuni*,   *fapte,*   *granularitate*;
      * schemele de date de bază   *model multidimensional*;
      * ce sunt fapte   *aditive*,   *semi-aditive*   și   *fapte non-aditive*;
      * ce sunt *dimensiuni ce se schimbă incet,* *măsuri ce se* *schimbă rapid,*  *dimensiuni degenerate*;
      * ce prezintă *„ierarhie* *de măsurare“*;
      * ce prezintă  *schema*  *„stea“*;
      * ce prezintă  *schema* *„fulg* *de* *nea“*;

***și învață :***

* + - * să modelați *tabele* de *fapt tranzacționale,tabele de fapte periodice instantanee,* *tabele factuale instantanee cumulative*;
      * să modelați  tabele   *dimensiuni*   și să dezvoltați   *ierarhii de dimensiuni*;
      * să utilizați trucurile de bază pentru rezolvarea relației "multi-la-multe" în   *modelul multidimensional*.

### CONCEPTE DE BAZĂ ALE METODEI DE MODELARE MULTIDIMENSIONALĂ

***Modelare multidimensională*   (*****Dimensional******modeling*) este mai ușor de înțeles decât ER - *modelare* .**

***Modelare multidimensională*** **este o metodă de modelare și vizualizare a datelor ca pe o multitudine de parametri numerici sau lingvistici, sau parametri (măsuri** *measures***),** **care descriu aspectele generale ale organizației.** De regulă, când  folosim *modelarea multidimensională principalul accent il punem pe datele numerice*, cum ar fi: numărul de vânzări, soldul,   *profit,*   *greutate*, *sau pe obiecte* care pot fi numărate, cum ar fi: articole, brevete, cărți ș.a.

*Modelare multidimensională*   are multe părți comune cu metoda de modelare " *entitate-relație*" pentru un model relațional, dar se deosebește prin scopurile ce le are.

Astfel, ***modelul relațional*** (pentru BD operatională) se concentrează pe integritatea datelor și eficiența întroducerii datelor.

***Model multidimensional***   (*Dimensional*   *model* ) este orientat în primul *rînd* pentru efectuarea interogărilor complexe la *BD*.

Metodă  de ***modelare multidimensională***   utilizează următoarele noțiuni: ***fapte, atribute*, *dimensiuni*,**   **parametri**   (metrici, măsuri)   **ierarhie,**   ***granularitate****.*

***Faptă***   (*fact*) - e**ste un set de elemente de date associate/interlegate, care conțin metrici/măsuri și date descriptive**. Fiecare    *faptă*   de obicei reprezintă un  *element de date,* care descrie numeric  *activitatea*   organizației, o tranzacție de afaceri sau un eveniment, care poate fi folosit pentru a analiza activitățile organizației sau procesele de afaceri.

În DWH   *faptele*  sunt stocate în tabelele din **BD relațională**. De exemplu:   *costul*  bunului produs/mărfii, *cantitatea* de mărfuri etc.

***Atribut***   (Attribute)   **- este o descriere a caracteristicilor obiectului real al domeniului** de studiu. De regulă   *atributul*   conține *valorile* cunoscute anterior  ce caracterizarea  o *faptă* . Uzual   *atributele* sunt   reprezentate de câmpuri de text cu valori discrete. De exemplu, caracteristicile ambalajului mărfurilor, mirosul mărfurilor ș.a.

***Dimensiunea***   (*dimension*) **- este o interpretare a faptei dintr-un anumit punct de vedere în lumea reală**.   *Dimensiunile,* asemeni    *atributelor*  conțin valori textuale, care sunt strâns legate între ele după sens.

De regulă *dimensiunile*   sunt reprezentate ca axe ale spațiului multidimensional, ale cărui puncte sunt asociate cu *faptele*.

În *modelul multidimensional*   fiecare   *faptă* este legată cu una mai multe axe/dimensiuni.   *Dimensiunile* de regulă reprezintă, de obicei valori non-numerice,   *variabile lingvistice*, cum ar fi filialele unei organizații, angajații unei organizații, clienții etc. De exemplu, atunci când analizăm vânzările de produse, fabricate sau vândute de o organizație, *printre dimensiuni*   **de obicei am putea remarca**  *timpul* , *cumparatori, vanzatori, locul,  locatia* de vânzare sau depozitarea mărfurilor.

*Dimensiunile*  sunt setate/definite prin listarea/enumărarea ***membrilor***/ **elementelor proprii** (members ).

***Membru/element al******dimensiunii*** **(dimensional** **member) – este un nume sau un identificator unic** **(variabilă lingvistică) utilizată pentru a determina poziția membrului/elementului**. De exemplu   *dimensiunea* "*timp*" poate conține următoarele membre/elemente ale dimensiunii:*" toate lunile "," trimestrele "," ani ".*

Adesea **membrul/elementul** *dimensiunii* se află **în legătură “parte- întreg” sau “copil-părinte”** , care vă permite să introducem pentru *dimensiuni*  una sau mai multe   *ierarhii*. Fiecare   *ierarhie*   poate avea mai multe    **ierarhiei**   ( *hierarchy* levels).  Fiecare membru/element al *dimensiunii*    ar trebui să aparțină doar unui singur nivel al ierarhizarii, generând astfel   *o secvență/tăieturi* de subseturi/submultimi ce nu se intersectează. Un exemplu ar putea fi   *ierarhia*   pe   *dimensiunea*   "*timp*": an, jumătate de an, trimestru, lună și zi.

**Membrul/elementul** *dimensiunii* “săptămână” poate să aparțină simultan la două luni, iată de ce pentru acest memebru/element al dimensiunii ar trebui definită o altă ierarhie.

**Opțiunea/parametrul, valoarea/metrica/măsura sau** *măsura***/indicatorul (measure)** **– este o caracteristică numerică a** *faptei***ce determină eficiența** **organizației** sau activității de business **din punctual de vedere al dimensiunii.**

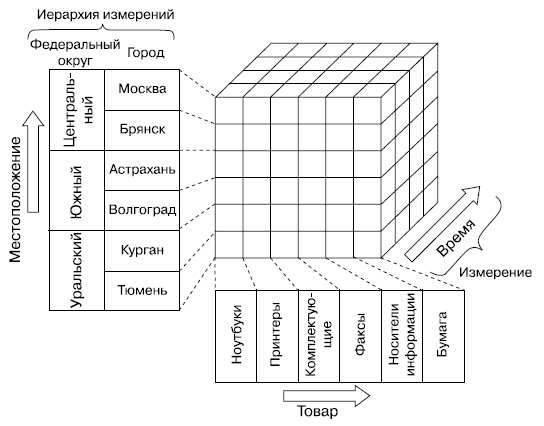
De regulă   *măsura*/*metrica*  conține valoarea prealabilă necunoscută a caracteristicii *faptă* . Valorile concrete ale *măsurilor* sunt descrise cu ajutorul variabilelor. De exemplu, fie că *măsura* este *expresia numerică*  a vânzărilor de bunuri exprimată în bani, numărul de unități de bunuri vândute etc.   *Măsura*/*metrica se* determină de o combinație de elemente a dimensiunii, și astfel reprezintă de fapt *o faptă* .

***Granulare***   (*Granularity*)   **- este nivelul de detalizare al datelor stocate în DWH** . De exemplu, vânzările zilnice.

### MODEL MULTIDIMENSIONAL

*Model multidimensional*   vizual poate fi prezentat folosind un *cub* (sau în cazul a mai mult de trei   *dimensiuni*   -  *hipercube*).

***Să considerăm un exemplu***. Fie că volumul de vânzări al unei organizații ce se ocupă cu comertul este o *funcție*  de următoarele variabile: "Produse", "Luna" și "Regiunea vânzărilor". Atunci, în calitate de *dimensiuni* vor fi "Bunurile", "*Timpul*"și "Locația"   În **Figura 1**. este prezentat *cubul de date multidimensional*    pentru a reprezenta această funcție.



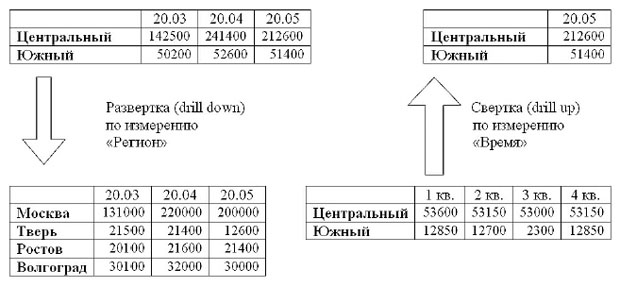
**Fig.** **1.**  Cub de date multidimensional

Pentru aceste   *dimensiuni*   pot fi definite următoarele ierarhii:

* *dimensiunea*  "Produse" - "Producător-Categorie-Produs";
* *dimensiunea*   "Locație" - " Districtul Federal -City-Shop";
* *dimensiunea*   "*Timp*"   - "An-trimestru-lună" sau "Săptămâna-zi".

*Modelare multidimensională*   este principala metodă de proiectare logică a DWH-urilor pentru aplicațiile OLAP . Pentru astfel de aplicații, este tipic să se efectueze operații de *pliere/împăturire/свертывания* și *desfășurare/развертывания* a datelor.

***Detaliere prin******desfășurare in jos***( *drill* down) și   prin ***pliere/împăturire în sus*** (*drill* up), operații de deplasare în jos și în sus pe nivelele *ierarhiilor dimensiunilor*. Atunci când efectuăm mișcarea *Detaliere prin desfășurare in jos* ( *drill* down),  *utilizatorul*  se mișcă de la nivelurile superioare la nivelele inferioare, care de obicei conțin date mai detaliate. Când efectuăm *pliere/împăturire în sus*(*drill* up), *utilizatorul*  se mișcă de la nivelurile inferioare ale ierarhiei la partea superioară, și astfel generalizînd informațiile pe fiecare nivel. În timpul executării acestor operațiuni   *calea*   navigării este determinată de *ierarhiile* *dimensiunilor*.



**Fig.**  **2.**  ***Detaliere prin******desfășurare in jos***( *drill* down) și   prin ***pliere/împăturire în sus*** (*drill* up),

Astfel, *ierarhiile* definesc modalități de *sumarizare/generalizare* a datelor, în cazul nostru a vânzărilor. De exemplu, **care a fost valoarea vânzărilor de bunuri HP LaserJet 1010 din categoria "Imprimante" a producătorului HP, volumul vînzărilor, în a treia săptămână a lunii noiembrie în orașul Bryansk din magazinul "Computing Machinery".**

Pentru ca   *Aplicația* să poată să utilizeze lista enumerată de *operații*   pentru prelucrarea datelor, este necesar să se plaseze datele în DWH într-un anumit mod, adică în formatul unui *model multidimensional* .

### FAPTE

În ceea ce privește interlegătura dintre  *dimensiuni*   și    *fapte,* vom menționa   că acestea din urmă pot fi împărțite în următoarele clase:

* ***fapte adiționale*** (Additive facts).   ***O faptă se numește aditivă* dacă are sens să o folosești *cu oricare altă*** *dimensiune* **pentru a** **efectua operații de sumare/generalizare cu scopul de a obține un rezultat semnificativ**. De exemplu, indicatorii numerici discreți ai activității de lucru, cum ar fi de exemplu numărul de vânzări, volumul vânzărilor, etc .;
* ***fapte*** ***semiaditive***   (Semiadditive facts*).****O faptă se numește*** ***semiadditivă* dacă are sens să o folosești împreună *cu unele****dimensiuni*   **pentru a** **efectua operații de sumare/generalizare cu scopul de a obține un rezultat semnificativ**. De exemplu, indicatorii numerici de intensitate, cum ar fi soldul contului, nivelul stocurilor dintr-un depozit etc.
* ***fapte non-aditive***    (Non-additive facts).   ***O faptă se numește non-additivă* dacă nu are sens să o folosiți împreună cu oricare dintre** *dimensiuni*   **pentru a** **efectua operații de sumare/generalizare cu scopul de a obține un rezultat semnificativ**.  De exemplu, *dimensiunea*   temperatura din cameră ;
* ***măsuri numerice a******intensității***   (Numerical Measures of *Intensity*).   ***O faptă se numeste masura numerică a intensitatii* daca ea, fiind**   *non-additivă*  **după** *timp***,** **permite agregarea și sumarea/generalizarea pe un anumit număr de perioade de timp**. De exemplu, restul pe cont.

Să precăutăm cîteva exemple de  *fapte non-aditive*.   *Faptele,* ce reprezintă atributele textuale, sunt non-aditive. Este evident ca să sumarizăm/rezumăm astfel  de *fapte* nu are sens. Pentru astfel de *fapte* este logic să numărăm numărul lor în tabelul *de fapte* (putem utiliza functii de tip *count()*).

*Prețul unei unități de marfă/produs* este de asemenea o faptă *non-aditiv*ă. Nu are sens să sumarizăm/ rezumăm prețul unei prize electrice sau sinecostul unei unități de produs.

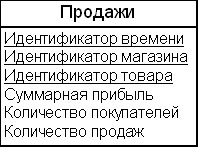
Cu toate acestea, dacă se depozitează produsul, costului unei unități de marfă/produs cu cantitatea de bunuri (marfă/produs) vândute, această valoare va fi deja  o *faptă aditivă*.

*Procente și raportul anumitor mărimi* (de exemplu *profitul* brut) sunt *fapte non-aditive*. Pot fi stocate ca parametri separați:   *numărătorul*   și *numitorul* relației, atunci când suma lor separată are sens. În acest caz aceste mărimi sunt déjà *fapte aditive*.

Unele mărimi cum ar fi sumele medii statistice, de exemplu, temperatura medie pe zi, sunt *fapte non-aditive*. Suma temperaturilor medii zilnice pentru o săptămână nu are sens.

Vom ilustra noțiunea de *faptă aditivă* pe un exemplu concret. Fie entitatea tabelei *fapte*   "Vânzări", **Figura.3,** cu   *atributele*   cheiei primare " *Identificatorul de* timp", "*identificatorul de* produs","*identificatorul de* magazin/shop" are următoarele   *fapte*  (valori/metrici/măsuri): "Numărul de cumpărători", "Total *profit* " și "Numărul de vânzări".

Deoarece sumarea/generalizarea poate fi efectuată pentru valorile /metricile/măsurile numerice ale entității "Număr de vânzări"și "Total *profit*" pe toate *dimensiunile*, aceste  *fapte*   sunt   *aditive,* așa cum se arată în exemplul de mai jos.



**Fig.** **3.**  **Un exemplu de entitate "Vânzări" cu fapte *aditive* și *semi-aditive***

**Exemplul 1.  *Faptele aditive*  pot fi sumate pe toate   *dimensiunile***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **pe *dimensiunea* "Timp":** | | | | | |
| **data** | **bunuri** | **Magazinul** | **Numărul de vânzări** | **Numărul**  **Cumpărători** | **Profitul sumar** |
| 23.01.2009 | CD- ul | Calculatoare | 10 | 10 | 1500 |
| 24.01.2009 | CD- ul | Calculatoare | 35 | 30 | 5250 |
| 25.01.2009 | CD- ul | Calculatoare | 20 | 15 | 3000 |
|  |  |  | **65** | **55** | **9750** |
| **pe *dimensiunea* " Produs ":** | | | | | |
| **data** | **bunuri** | **Magazinul** | **Numărul de vânzări** | **Numărul**  **Cumpărători** | **Profitul sumar** |
| 23.01.2009 | CD- ul | Calculatoare | 10 | 6 | 1500 |
| 23.01.2009 | imprimantă | Calculatoare | 1 | 1 | 5000 |
| 23.01.2009 | scaner | Calculatoare | 2 | 2 | 3000 |
|  |  |  | **13** |  | **9500** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **pe *dimensiunea* " Magazin ":** | | | | | | | |
| **data** | **bunuri** | **Magazinul** | **Numărul de vânzări** | | **Numărul**  **Cumpărători** | **Profitul sumar** | | |
| 23.01.2009 | CD- ul | Calculatoare | | 10 | 10 | | 1500 |
| 23.01.2009 | CD- ul | Imprimante | | 10 | 5 | | 1500 |
| 23.01.2009 | CD- ul | Echipamente de birou | | 20 | 7 | | 3000 |
|  |  |  | | **40** | **22** | | **6000** |

Vom ilustra conceptul  de *fapte* *semiadditive*   printr-un exemplu. Fie entitatea tabelei *fapte* "Vânzări" **Figura 3.** După cum se poate vede din Exemplul 1, *dimensiunea*   "Produs", sumarea valoarea/metrica/măsura "Numărul de cumpărători" nu a fost efectuată, iar pe *dimensiunile*  "*Timp*" și "Shop" sumarea a fost realizată. Acest    *fapt*   este   *semiadditiv*   în legătură dimensiunea   „Produse“.

**Exemplul 2** .   *Fapte* *semiadiante*   Nu are sens să rezumîm/sumam pentru unele  *dimensiuni*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sumarul valorii/metricii/măsurii "Numărul de cumpărători" pe dimensiunea "Produs"** | | | | | |
| **data** | **bunuri** | **Magazinul** | **Numărul de vânzări** | **Numărul**  **Cumpărători** | **Total**  **profit** |
| 23.01.2009 | hârtie pentru faxuri | Calculatoare | 10 | 6 | 1000 |
| 23.01.2009 | hârtie pentru imprimantă | Calculatoare | 12 | 7 | 1320 |
|  |  |  |  | **13** |  |

În Exemplul 2 este aratat   *tabelul*  în care a fost efectuată sumarea *valorii/ metricii/măsurii* "Numărul de cumpărători"   *pe dimensiunea*   „Produse“. Să punem următoarea întrebare: *"Câți clienți au cumpărat fie hârtie pentru imprimantă, fie hârtie pentru fax?"*. Răspuns: undeva între 5 și 13. Incapacitatea de a împărți numărul de cumpărători între bunuri pentru entitatea dată face acest lucru   *să* fie o *faptă*   *semiadditivă*    pentru   *dimensiunea* „Produs“.

### CHEI IN TABELUL DE FAPTE

*Cheia*  primară în tabelul  *de fapte*   este de regulă o *cheia*  primară compusă. Ea constă din   *seturi*   de *chei*  externe    care servesc în calitate de *chei*  primare a *dimensiunilor* legate cu tabelul de  *fapte*.

Trebuie de remarcat faptul că uneori mai multe combinații de *chei* externe reprezintă o *cheie* primară unicală/*unică* pentru tabele   *de fapte* . Pe de altă parte, nu chiar totdeauna combinația dintre toate *cheile* externe  *ale dimensiunilor* garantează unicalitatea/*unicitatea* *cheiei*  primare  a  tabelei   *de fapte*   (mai ales în cazul în care în tabelul   *de fapte*  există   *dimensiuni degenerate,* despre care vedeti.  mai jos).

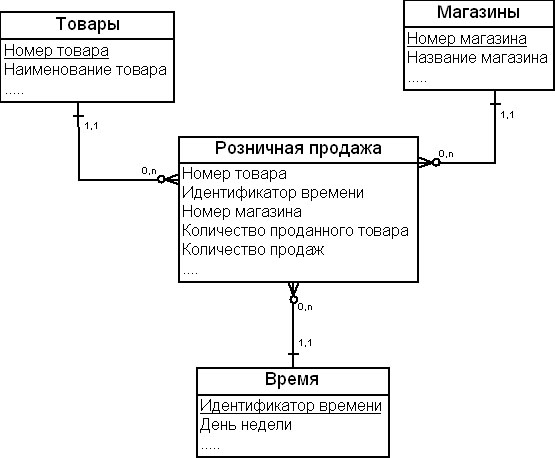
Să analizăm în detaliu problema unicalității/*unicității* *cheiei*  primare  a  tabelei   *de fapte* .

***Faptele,* de regulă sunt granulare**.   *Granularitatea*    *faptelor*   definește semnificația semantică a sensului  *faptă*   în ceea ce privește nivelul de detaliere, legată/asociată  de *faptă* informație. De exemplu, *vânzările totale* pentru un *magazine/shop concret,* *într-o anumită zi,* *pentru un tip de produs dat*.

***Deseori   granularitatea  se determină la nivelul unei business-operațiuni*** . De exemplu,   *vânzarea*   de mărfuri. Uneori este necesar să salvăm informația despre vânzarea acestui *produs pentru ziua respectivă* (aici există deja un anumit nivel de agregare a datelor).

Unicitatea/unicalitatea  *cheiei* primare pentru  tabela   *de fapte*  este *garantată/ determinată* de definirea  *granularității*   *faptelor*, și nu a condiției de unicitatea/unicalitate a *cheilor*  externe a tabelelor *dimensiuni,* dacă  *granularitatea faptelor*   nu este definită cum trebuie.

Să precăutăm în continuare tabelul   *de fapte*   din diagrama  **Figura 4.** *Granularitatea faptelor*  in ea alcătuieste o zi.   *Faptele*  se determină pentru acestă tabelă ca sumă tuturor tranzacțiilor pentru această zi.



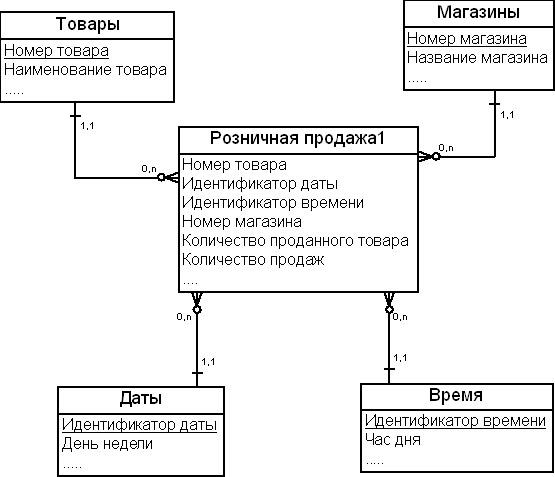
**Fig.** **9.4.**  Exemplu de Schema de date pentru a ilustra unicitatea/unicalitatea cheiei primare a tabelei de fapte, granularitate faptelor – 24 ore /1 sutca/

Să presupunem că în   *dimensiunea*   "Produs" există două tipuri de plăci - C1 și C2.   ***Dimensiunea   "Data"*** conține un singur rind pentru o zi. Să presupunem că plăcile de tip С1 au fost vândute la 1 martie 2009 la ora 9 dimineața în cantitate de **4 bucăți** și seara înainte de închiderea magazinului în cantitate de **5 bucăți.** Deoarece   *granularitatea*  *faptelor* este o zi, atunci în tabela   *de fapte*   "Vânzările cu amănuntul" pentru fiecare produs vândut, în fiecare magazin, vor fi introduse doar cite un singur rind/inscriere pe zi. Deci, pentru 1 martie 2009 în  ***câmpul*   "Numărul de vânzări" va fi numărul 9.**

Dacă noi vom defini *cheia*  primară pentru tabela  *de fapte*   ca concatenări de *chei*   primare    *a dimensiunilor*  {***Număr articol,   identificator   date, număr magazine/shop***}, atunci această *cheie*  primară    va fi unică/unicală, deoarece *fapta*   se va exprima printr-o singură linie/inscriere, introdusă pentru fiecare produs, care este vândut în fiecare magazin într-o anumită zi.

**Unicitatea/ unicalitatea *cheiei*  primare este  garantată de definirea   *granularității*   *faptelor*** .

Să presupunem că conducerea unei Companii comerciale are nevoie de rapoarte de vânzări a bunurilor **în fiecare oră**. Atunci putem să folosim schema prezentată  din  **Figura 5.**



**Fig.** **5.**  Exemplu de Schema de date pentru a ilustra unicitatea/unicalitatea cheiei primare a tabelului de fapte : *granularitatea faptei - o oră*

Să presupunem că, în decursul zilei de 12 martie 2009, plăcile de tip С1 au fost vândute după cum urmează Tabelul 1.

|  |  |
| --- | --- |
| Tabelul 1. Informații privind vânzările de plăci | |
| **timp** | **Numărul de persoane** |
| 8-00 | 5 |
| 8-30 | 5 |
| 11-00 | 2 |
| 16-00 | 3 |
| 18-00 | 2 |

**Deoarece   *granularitatea*   tabelei   *de fapte***  este exprimată prin prezența unui rând/inscrieri pentru fiecare produs, **vândut în fiecare magazin în fiecare oră,** în table vor **fi 4 rinduri,** prin care se stabilește vânzarea plăcilor de tip C1 pentru 2 martie 2009..

Dacă noi vom defini *cheia* primară    a  tabelei   *de fapte*   ca concatenări ale *cheelor*  primare  *a dimensiunilor* {***Număr articol,  identificator   de timp   identificator   date, număr magazine/shop},*** atunci această *cheie* primară va fi unică/unicală.

Astfel,   *granularitatea*   *faptei* garantează/determină unicitatea/unicalitatea *cheiei* primare  compuse   a tabelei   *de fapte*.

### TABELE DE FAPTE

***Faptele*  în   *modelul multidimensional*   este obișnuit să le prezentăm ca un table denumit *fapte*.**

**În modelul logic** "*entitate-relație*"   *tabelul*    *faptelor*  este prezentat de o entitate,  *atributele*   carei sunt   *faptele*   (valori/metrici/masuri sau descrieri) și *cheia* compusă ce asigură legătura tabelei  *de fapte*   cu tabelele   *de dimensiuni*   prin legături de tipul "*unu-la-mulți*".

**Tabelele  *de fapte*  sunt  împărțite în trei categorii principale, în funcție de nivelul de detaliu *al faptelor*  (*granularitatea*).**

* ***Tabelul tranzacțional*** ***al faptelor***. Într-un astfel de tabel  *de fapte*  sunt salvate   *fapte* ***care fixează anumite evenimente*** (tranzacții). Acestea sunt  *fapte* ce descriu de obicei fiecare in parte un eveniment din activitatea de business. De exemplu, *vânzări de bunuri.*
* ***Tabelul imaginilor periodice de moment al faptelor.***   Într- un astfel de tabel sunt colectate *Fapte,* ***care stabilesc starea actual/curentă a unei anumite linii de afaceri.*** Acestea sunt *Fapte,* care descriu ***starea curentă*** a unei anumite linii de activitate a busienssului ***pentru orice combinație de valori a dimensiunilor pentru perioadă  dată de timp***. De exemplu,  vânzările organizației la o anumită data ( zilnic ).
* ***Tabelul de fapte al imaginilor instantanee cumulative***   Într- un astfel de tabel sunt colectate *faptele care* fixează ***o anumită stare finală*** a unei anumite linii de afaceri **în momentul prezent *de timp***. Acestea sunt *fapte*, care descriu totalurile intermediare ale activității companiei pe o anumită directie de activitate a businesului pentru orice combinație a valorilor *dimensiunilor* ***pentru o anumită dată***.

***CARACTERISTICILE DE BAZĂ ALE TABELEI DE FAPTE SUNT:***

* 1. **Tabelul *faptelor* conține parametri numerici** (valori/metrici/măsuri).
  2. **Fiecare tabel *de fapte* are o *cheie* compusă**, ce constă din *cheile* primare ale tabelelor de *dimensiuni*. *Cheia* primară a tabelului de *dimensiuni* este *cheia externă* în tabelul de fapte.
  3. ***Tabela de fapte are, de regulă, un număr mic de cimpuri, nu mai mult de 20***.
  4. ***Datele din tabelul de fapte au următoarele proprietăți:***
* ***parametrii numerici*** sunt folosiți *pentru agregare și sumarizare*;
* ***valorile datelor trebuie să aibă proprietăți de aditivitate sau semiaditivitate***în raport cu *dimensiunile*, pentru a le putea rezuma/sumariza;
* ***toate datele tabelului de fapte trebuie să fie identificate în mod unic*** prin *cheile* tabelelor de *dimensiuni* pentru a le oferi acces prin tabelele de *dimensiuni*.

**Astfel, tabelul *de fapte* poate fi împărțit în două părți.** **Prima parte** constă din *cheile primare* ale dimensiunilor, **iar a doua parte** - din parametrii numerici ce depend funcțional de *cheile* tabelelor de *dimensiuni*.

Deseori, tabelă de *fapte* are o ***cheie primară surogată***, generată de sistem, de care, prin definiție, toți parametrii depind ***funcțional complet***.

Un exemplu de *tabel de fapte* este prezentat în **Fig. 3.**

**Notă:** Cheia primara “surogat” este folosita pentru a inlocui cheia primara compusa din mai multe coloane. Cheia primara surogat este o coloana cu valori generate in ordine secventiala, folosind un dispozitiv de incrementare. In Access este utilizat in acest scop tipul de camp autonumeric  (autonumber). **Cheile surogat** sunt numite chei fara semnificatie, deoarece din valoarea acestora nu sunt obtiunute informatii privind inregistrarile (liniile) pe care le identifica.

**DIMENSIUNILE**

**Principalele caracteristici ale tabelului de *dimensiuni* sunt următoarele.**

1. Tabelele de *dimensiuni* conțin ***date privind detaliile*** despre *fapte.*

2. Tabelele de *dimensiuni* conțin ***informații descriptive*** despre valorile numerice din tabelul de *fapte*, adică ele conțin *atribute ale faptelor*.

3. De regulă, *tabelele de dimensiuni denormalizate*, conțin ***un număr mare de câmpuri.***

4. Tabelele de *dimensiuni* conțin ***de obicei mai puține rânduri decât tabelele de fapte.***

5. ***Atributele*** tabelelor de *dimensiuni* ***sunt utilizate la vizualizarea datelor în rapoarte și interogări.***

În **Figura 6** este prezentat un exemplu al tabelului de *dimensiuni* "Timp". "*Identificatorul* de timp" este *cheia primară* a tabelului de *dimensiuni*. Câmpurile rămase *sunt atribute* ale parametrilor din tabela de fapte, dependente de *timp*.



**Fig. 6.** Tabel de *dimensiuni*

De regulă, *tabela de dimensiuni* are o *cheie primară* și *atribute* care descriu *faptele* din punctul de vedere al unei anumite directii de activități ale organizației.

**PRINCIPALELE SCHEME ALE MODELULUI MULTIDIMENSIONAL**

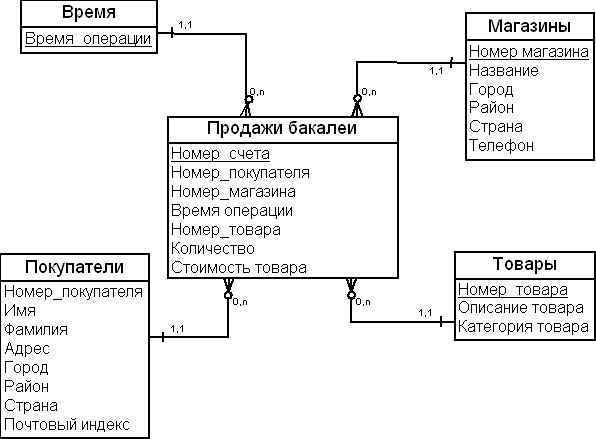
***Există mai multe scheme de modelare a datelor multidimensionale.*** Două dintre ele sunt considerate de bază: ***schema stea*** (*star schema*) și schema ***fulg de zăpadă*** (*snowflake schema*). În cazuri mai complexe, se folosesc așa-numitele scheme "***multi-stea***" sau o ***schema cu mai multe tabele de fapte***.

Schema “***stea”*** **are un *tabel de fapte* și mai multe *tabele de dimensiuni*.** **Tabelele de dimensiuni sunt denormalizate.**

Schema “***fulg de zăpadă”*** **are un *tabel cu fapte* și mai multe *tabele de dimensiuni normalizate*.**

Schema “***stea”*** are aceleași elemente ca și diagramele ***entitate-relație***. Acestea sunt ***entități, atribute, chei primare și externe/străine***, interconexiuni, legături cardinale.

În **Figura 7** este prezintat un exemplu de schemă **“*stea”,*** creată pentru a ține evidența vânzărilor produselor alimentare. Tabelul “***de fapte”*** de “Vânzări de produse alimentare” (evidenta vânzărilor de produse alimentare de către o societate comercială) are ***cheia primară*** "Număr cont", 4 ***chei externe*** (egal cu numărul de *dimensiuni*) și doi parametri: "Cantitate" (bunuri vândute) și "Costul bunurilor".



**Fig. 7.** Schema “Stea”

*Dimensiunea* “*Ora*”(Timpul/Vremea) este unul dintre elementele critice ale modelului DWH.

**Dacă datele din interogările sistemelor OLTP se concentrează asupra momentului actual/current de timp,** **atunci**

**in sistemele de suport decizional (sistemele DSS) OLAP, pentru care DWH -urile sunt proiectate și create,** **interogările se concentrează asupra sarcinilor de *analiză a datelor*, și anume,** ***modul în care datele s-au schimbat în diferite perioade de timp*.**

De exemplu, care a fost volumul vânzărilor unei societăți comerciale pentru **ultimul trimestru**, l**ună** sau **care sunt tendințele privind achizițiile de bunuri în ultimul trimestru**?

***Dimensiunea*** “Magazin” permite să grupam operațiunile de vânzări de către magazine lu]nd in considerare locația lor geographica.

***Dimensiunea*** "Mărfuri/Produse" permite să analizam schema tipică de achiziționare de bunuri și să răspundem la întrebarea, ce produse sunt de obicei cumpărate în același timp de către cumpărători.

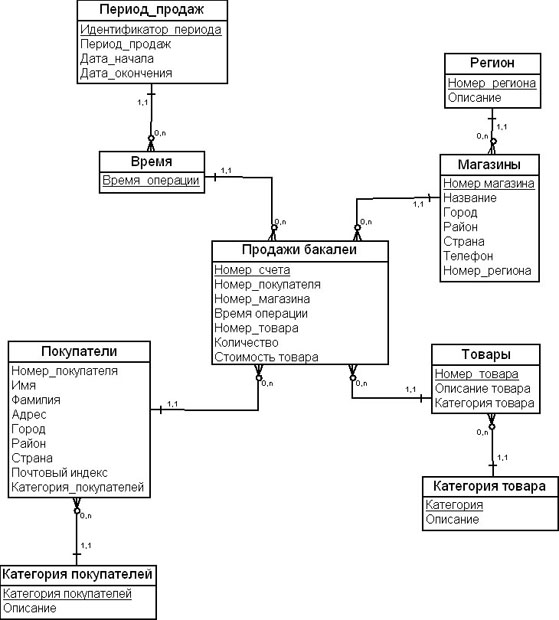
***Dimensiunea*** "Cumpărători" permite să analizam cumpărăturile luind în considerare frecvența lor, locației geografice și cantității.

**TABELELE DE *DIMENSIUNI* SUNT UN FEL DE *GHID* ATUNCI CÂND SELECTAȚI RÂNDURI/ÎNSCRIERI DINTR-UN TABEL DE *FAPTE*.**

Tabelele ***de fapte*** au un număr mare de rînduri/înscrieri/înregistrări. Tabelele de ***dimensiuni*** au, de obicei, un număr mult mai mic de rînduri/înscrieri/înregistrări.

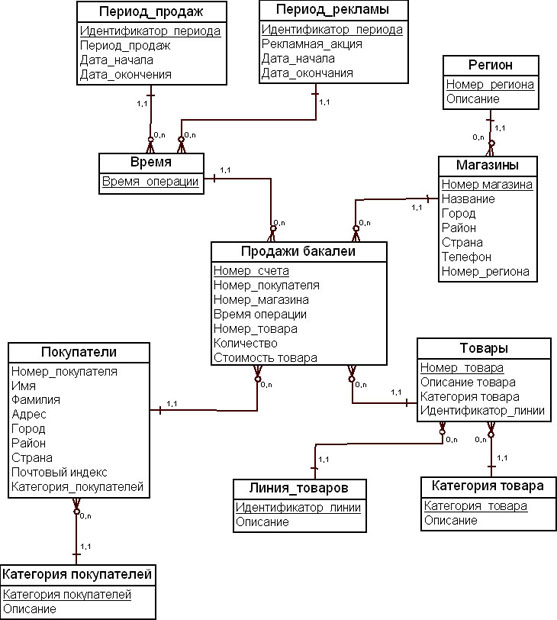
Unul dintre principalele avantaje ale utilizării unei astfel de scheme este că performanța operației SGBD cu a uuni *tabelul de fapte* și a unor astfel de tabele de conexiuni (de dimensiuni) va fi aproape de o conexiune optimal (în sensul accesării la date).

***Modelul de fulg de zăpadă*** **Figura 8** ***adaugă o ierarhie în tabelele de dimensiuni.*** De exemplu, ***dimensiunea* "Regiune"** grupează magazinele în funcție de geografia regiunii, ***dimensiunea* "Categoria de Produs"** grupează produsele în categorii, ***dimensiunea* "Categoria de Clienți"** grupează clienții în categorii, iar ***dimensiunea* "Perioadă vânzări"** grupează vânzările după perioade de *timp*.

[](https://www.intuit.ru/EDI/08_01_19_2/1546899581-11707/tutorial/632/objects/9/files/07_08.jpg)

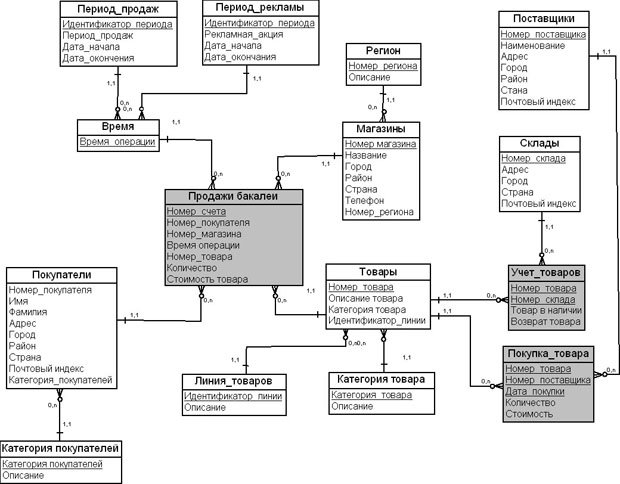
**Fig. 8. Schema "Fulg de zapada"**

Astfel, **folosirea ierarhiilor** **transformă schema** *"Stea"* în schema *"Fulg de zăpadă"*. In cele din urmă noi putem extinde și schema *"Fulg de zăpadă"* prin includerea în ea de noi ierahii, după cum este arătat în schema din **Figura 9.**

[](https://www.intuit.ru/EDI/08_01_19_2/1546899581-11707/tutorial/632/objects/9/files/07_09.jpg)

**Fig. 9.** **Adăugarea de noi ierarhii în schema “Fulg de zăpadă”**

În **Figura 10** este prezentat un exemplu de schemă cu mai multe ***tabele de fapte*** - Schema de „***constelație a faptelor***” (*Fact* Constellation *Schema*). **Diagrama are trei tabele *de fapte*:** „**Vânzări de produse alimentare”, „Evidenta\_produse” și „Cumparare\_produse”**, care au atât *dimensiuni* comune ("Produse"), cât și exclusive (de exemplu, *dimensiunea* „Depozite” pentru tabela *de fapte* „Evidenta Produselor”).

[](https://www.intuit.ru/EDI/08_01_19_2/1546899581-11707/tutorial/632/objects/9/files/07_10.jpg)

**Fig.**  **10. Schema “Constelatie de fapte”**

În cele ce urmează, vom examina mai atent aspectele ce tin de modelarea tabelelor de *dimensiuni*.

**MODELAREA TABELELOR DE DIMENSIUNI PRIN TABELELE DE FAPTE**

Când modelează tabelele   *de fapte*   designer-ul DWH în *primul rînd* se bazează pe luarea în considerare a unei astfel de caracteristici, cum ar fi gradul de detaliere (*granularitate*) *a faptelor*.

Mai sus noi am precăutat tabelele *de fapte*, care conțin parametri (valori/metrici/măsuri) la cel mai jos nivel de detaliere (*granularitate*), cum ar fi nivel *operațiunilor* de afacerialeorganizației sau *tranzacțiele instantanee/de moment/ad hoc*.

Tranzacțiele instantanee/de moment/ad hoc oferă posibilitatea de a înregistra o descriere *a faptelor*  pentru un moment dat *de timp*. Acestea sunt ***tabele de fapte tranzacționale*** .Un exemplu de astfel de tabel este afișat în   [Fig.](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D2%23image.9.3#image.9.3) [9.3](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D2%23image.9.3#image.9.3) .

*Faptele* „Profitul total", "Numărul de vânzări" și "Numărul de cumpărători" sunt înregistrate în tabela *de fapte*   la nivelul   *operațiunilor de* vânzare de bunuri (tranzacție instantanee).

Atragem atenția asupra faptului că astfel de tabele *de fapte*   pot conține *atribute*  text  și *fapte* numeric non-aditive.

Cu fiecare nivel al  *granularității* poate fi conectat un ***agregat***. **Agregatele sunt sume ale valorilor parametrilor sau ale funcțiilor statistice ale valorilor parametrilor, preluate la un anumit nivel de detaliere.**  (*granularitate*). Agregatele pe unele si aceleasi fapte, pot fi definite la unul sau mai multe nivele de detaliu.

De obicei, în DWH sunt utilizate două tipuri de tabele aggregate *de fapte* după cum a fost relatat mai sus:

1. cu puterea de granularitate la nivelul imaginii instantanee periodice a datelor, reprezentând intervalul *de timp*  de durata dată  (*tabelul de fapte al imaginii instantanee periodice*);
2. cu puterea de granularitate la nivelul imaginii acumulatoare, care reprezintă întreaga istorie   *de fapte*   (date istorice) de la momentul dat și până în momentul prezent *de timp*   (*tabelul de fapte cumulative al imaginii instantanee*).

Agregatele au un mare *sens*   pentru performanța interogărilor.   *Interogarea* va fi executată mult mai repede pe 10 rânduri apriori calculate ale tabelului de  *fapte,* decît lazece mii de rînduri de un nivel mai jos de detaliere.

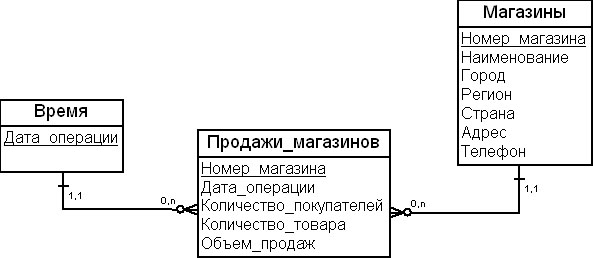
De exemplu, pentru schema de pe   [Fig.](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D3%23image.9.6#image.9.6) [9.6](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D3%23image.9.6#image.9.6)   se poate crea un agregat "de vânzare zilnică" pentru magazine sau regiuni, putem construi un agregat "vânzări săptămânale" pe categorii de produse.

**Tabel de fapte** **agregate** **(Aggregate fact table) se numește tabelul de fapte, care conține agregatele unor fapte ale modelului**.

Designer-ul adaugă tabelele *de fapte*  agregate în   *model multidimensional*  de date și stabilește legăturile ei cu tabelele *dimensiuni*.

Vom construi un tabel *de fapte* aggregate "Vânzările magazinelor" care conțin agregate   *de fapte*  a vânzărilor zilnice ale magazinului așa cum se arată în   [Fig.](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D4%23image.9.11#image.9.11) [9.11](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D4%23image.9.11#image.9.11) . Acest   *tabel*  conține totalurile sumelor pentru vânzările zilnice ale fiecărui magazin.

* Cîmpul “Numărul de cumpărători” indică, cât de multi cumpărători au efectuat cumpărăturile.
* Cîmpul "Cantitatea de bunuri" indică, câte unități de mărfuri au fost vândute.
* Câmpul "Volumul de vânzări" indică cantitatea de bunuri vândute pe zi.

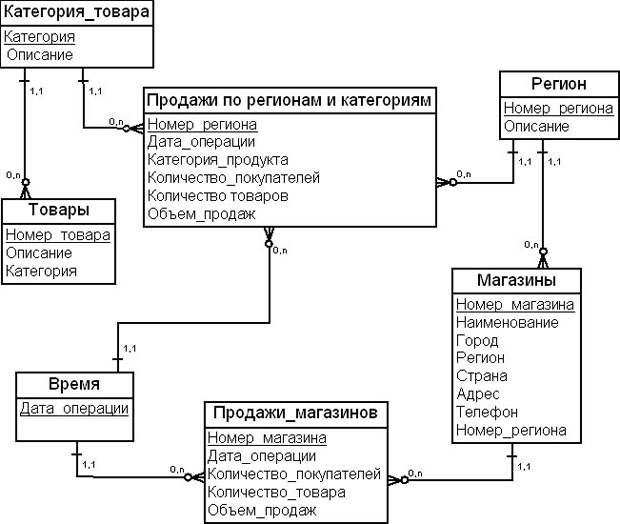


**Fig.** **9.11.**  Tabelul agregate de fapte "Vânzările magazinelor"

Acesta este un exemplu de *tabele de fapte ale imaginii instantanee periodice*.

În baza acestui tabel de agregate *de fapte* pot fi construite agregate *derivate*, cum ar fi numărul mediu de achiziții efectuate de un cumpărător, sau suma medie, cheltuită de un cumpărător pentru cumpărarea de bunuri.

Designerul poate proiecta un alt tabel agregat de *fapte*, care conține date privind vânzările zilnice pe regiuni și categorii de produse, după cum se arată în   [Fig.](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D4%23image.9.12#image.9.12) [9.12](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D4%23image.9.12#image.9.12) .

[](https://www.intuit.ru/EDI/08_01_19_2/1546899581-11707/tutorial/632/objects/9/files/07_12.jpg)

**Fig.** **9.12.**  Tabel de agregate de fapte "Vânzările magazinelor"

Pot fi adăugate un număr *arbitrar*  de tabele agregate *de fapte* în   *model multidimensional*  al DWH- ul. De menționat, că tabelele agregate de *fapte* prezintă prin sine răspunsuri apriori pregătite la întrebări cunoscute. Nevoile neplanificate de interogări la aggregate, de regulă, nu sunt în general suportate.

Pentru tabelul de *fapte,* legate cu 10  tabele de *dimensiuni* , există mai mult de trei milioane de tabele agregate *de fapte*. Ce tabele de agregăte *de fapte*  trebuie să fie construite în modelul de date, depinde de *obiectivele proiectului*  DWH și soluții de design a celui ce proiectează DWH.

La determinarea agregatelor este util să se folosească principiul Pareto: 20% dintre posibilii candidați pentru agregare vor fi într-adevăr solicitați de utilizatori.

*Tabelul de fapte cumulative al imaginii instantanee* agreghează in rândul tabelului *de fapte* date pentru tot *timpul* de  existență a oricărui eveniment. De exemplu *tabelul de fapte* al imaginilor cumulative pot stoca informații despre

cardurile de salarizare ale angajaților oricărei organizații. De regulă, aceste carduri sunt emise de bancă pentru o perioadă de la doi până la cinci ani, iar apoi se schimbă pe carduri noi.

În calitate de un astfel de tabel de aggregate, vom precauta compania, care se ocupă de angajarea în cîmpul de muncă. Să presupunem că în fiecare zi compania face publicitate în presă la mai mult de 10.000 de locuri de muncă. Conducerea companiei ar fi dorit să evalueze eficiența angajaților săi, care se ocupă de angajarea persoanelor în cîmpul muncii, și anume: câte  *timp*  durează procesul de angajare a candidatului, câte  *timp* cheltuite în așteptarea primirii *rezumatelor,* cât *timp* petrece *managerul*  companiei pentru angajarea candidatului care a solicitat serviciile companie de angajarei.

În acest tabel   *de fapte*   un rînd (faptă) va conține trei perioade *de timp* pentru fiecare manager și pentru fiecare post vacant. Exemplu de  schemă este prezentată în [Fig.](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D4%23image.9.13#image.9.13) [9.13](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D4%23image.9.13#image.9.13).

[](https://www.intuit.ru/EDI/08_01_19_2/1546899581-11707/tutorial/632/objects/9/files/07_13.jpg)

**Fig.** **9.13.**  Tabelul agregat de fapte "**Ocuparea forței de muncă**"

Acest tip de tabele agregate *de fapte*   este utilizat, de regulă, pentru procesele pe termen scurt și pentru procesele, bazate pe evenimente aleatorii, cum ar fi o tranzacție bancară.

[Tabel.](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D4%23table.9.2#table.9.2) [9.2](https://translate.google.com/translate?hl=ro&prev=_t&sl=ru&tl=ro&u=https://www.intuit.ru/studies/courses/599/455/lecture/10167%3Fpage%3D4%23table.9.2#table.9.2)   o comparație a principalelor tipuri de tabele *de fapte* .

|  |  |  |  |
| --- | --- | --- | --- |
| Tabelul 9.2. Compararea tipurilor de tabele de date | | | |
|  | ***tabelul tranzacțional***  ***de fapte*** | ***tabele de fapte ale imaginii instantanee periodice*** | ***Tabelul de fapte cumulative al imaginii instantanee*** |
| definirea *granularității*  tabelei de fapte | Un rând pentru fiecare tranzacție de afaceri | un singur rind pe perioadă | un singur rind pe întreaga perioadă a evenimentului finalizat |
| dimensiune | Se folosesc faptele la cel mai scăzut nivel de detaliere pe dimensiunea „data / ora“ | Se folosesc faptele la un nivel oarecare de agregare, pe dimensiunea „Data / Ora“ (după sfârșitul perioadei ) | Utilizați fapte din mai multe dimensiuni de dată / timp pentru a înregistra rezultatele la diferite puncte de control. |
| Numărul total de dimensiuni abordată | Mai mult, decât în ​​tabele de fapte de imagini periodice | Mai puțin decât  în *tabele de fapte tranzacționale* | Cele mai multe dimensiuni pentru tabelele de fapte |
| Fapte | Fapte legate de activitățile operationale | Fapte legate de activități periodice | Faptele ce sunt legate de activități care au un anumit timp de existență. |
| Actualizarea | nu sunt permise | nu sunt permise | permis |
| cardinalitatea  tabelelor de fapte | creștere repede | Creste mai lent decat in   *tabele de fapte tranzacționale* | Creste mai lent decat in   *tabele de fapte al imaginii instantanee periodice* |