Inteligenta afacerii
Cursul 2

Conf. Ramona Bologa,
ASE Bucuresti

Agenda

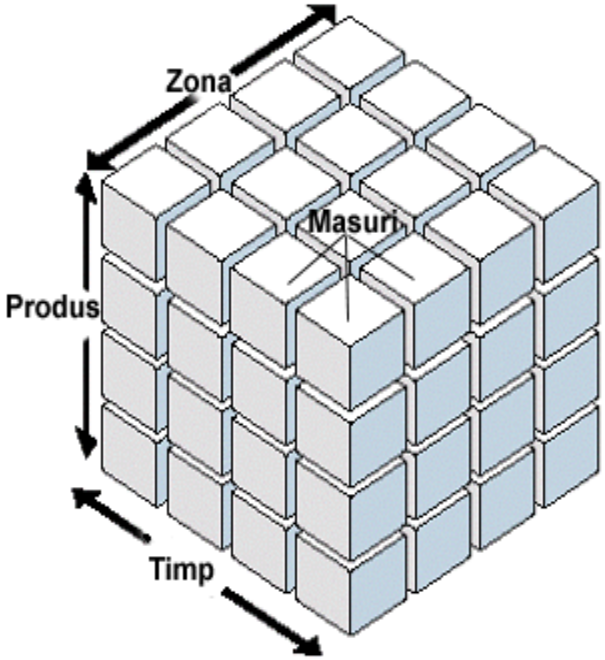
1. Arhitectura **depozitului de date**. Concepte de baza: cuburi, datamart, dimensiuni, fapte, agregari, granularitate.
2. Structura depozitului de date
   - Schema stea,
   - Schema fulg de zapada
   - Schema constelatie de fapte
3. Instrumente **ETL**
4. Concepte despre **datamining**

1. Arhitectura DW: modelul multidimensional

- Ralph Kimball – unul dintre arhitectii conceptului de depozit de date
- A elaborat o metodologie pentru proiectarea data mart-urilor, care conduce la o structurare a datelor intr-un model usor de folosit si foarte rapid
- Aplicarea metodologiei Kimball poarta numele de MODELARE MULTIDIMENSIONALA
- Procesul de proiectare are 4 pasi:
  1. Selectarea **domeniului de interes**.
  2. Declararea nivelului de **granularitate** a procesului
  3. **Alegerea dimensiunilor** care se pot aplica pentru fiecare linie din tabela de fapte si definirea atributelor
  4. **Identificarea faptelor numerice** care vor popula fiecare rand din tabela de fapte.

1. Arhitectura DW: modelul multidimensional

- permite vizualizarea datelor prin mai multe filtre sau **dimensiuni** in acelasi timp.



- Dimensiuni=coordonate= categorii de informaţie.

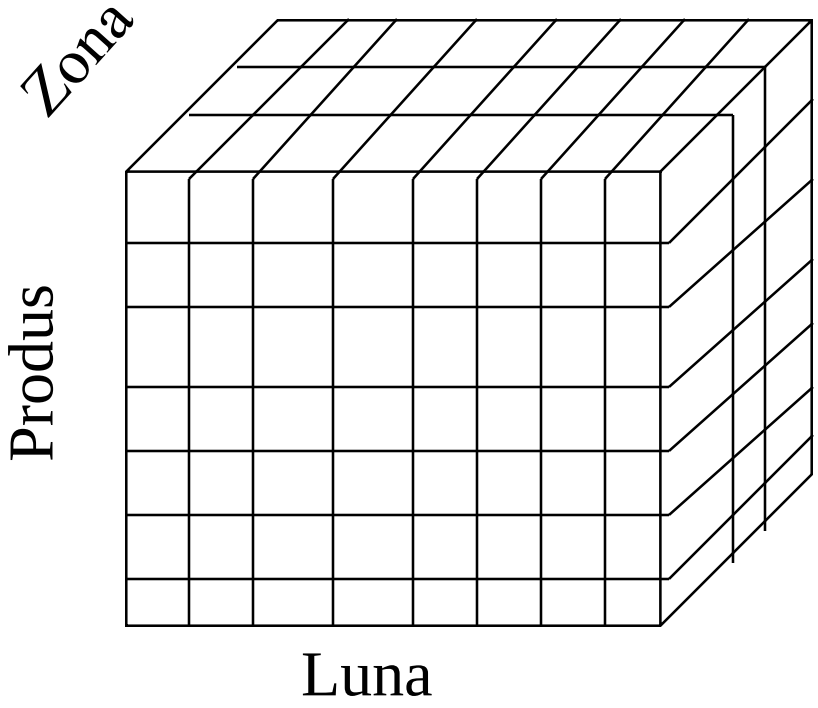
- De ex:
  - Care sunt vanzarile reale in comparatie cu cele previzionate pe ***zona***, pe ***vanzator***, pe ***produs***?
  - Care este profitabilitatea pe ***produs***, pe ***client***?

Obiecte DW

- **Tabelele de fapte** (masuri)
  - conţin **faptele** şi **cheile externe** către tabelele de dimensiuni.
  - de obicei *date numerice* - totalizate şi analizate pe diferite niveluri.
- **Tabele dimensiuni**
  - categorii de informatii care organizeaza datele
  - fiecare tabelă dimensiune are câte o **cheie principala**
  - câmpurile sunt de obicei *textuale* - **sursă pentru restricţii** şi pentru **rândurile din rapoarte**.
  - datele sunt de obicei colectate la nivelul cel mai de jos şi mai detaliat şi agregate pe nivelele superioare pentru analiză.
- **Atribut** - un nivel al unei dimensiuni, intr-o **IERARHIE**
- **Ierarhiile**
  - sunt structuri logice utilizate pentru ordonarea nivelelor de reprezentare a datelor.
  - definesc **caile de navigare** în interiorul datelor, permiţând detalierea graduală a datelor.

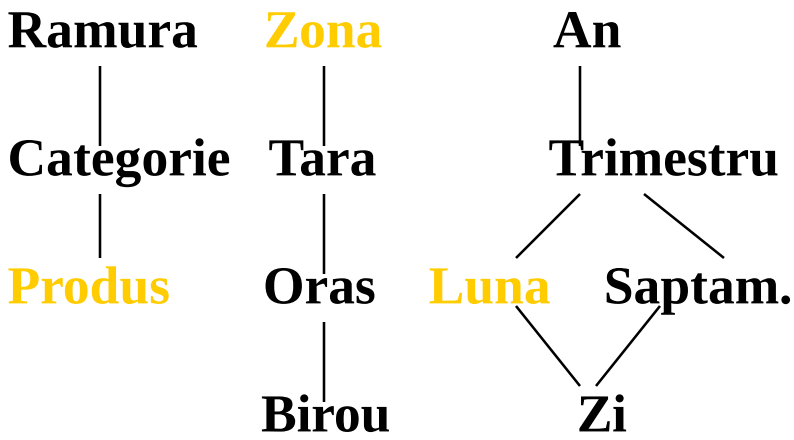
Date multidimesionale

- **Volumul vanzarilor** – in functie de produs, luna, si zona



Dimensiuni: Produs, Zona, Timp

Ierarhii:



Exemplu: Vanzari de fructe

Timp	Suma
Trim 1	16000
Trim 2	16000
Total Timp	**32000**

Piata	Suma
Brasov	8000
Sibiu	8000
Arad	8000
Iasi	8000
Total Piata	**32000**

Produs	Suma
Mere	8000
Cirese	8000
Struguri	8000
Pepeni	8000
Total Produs	**32000**

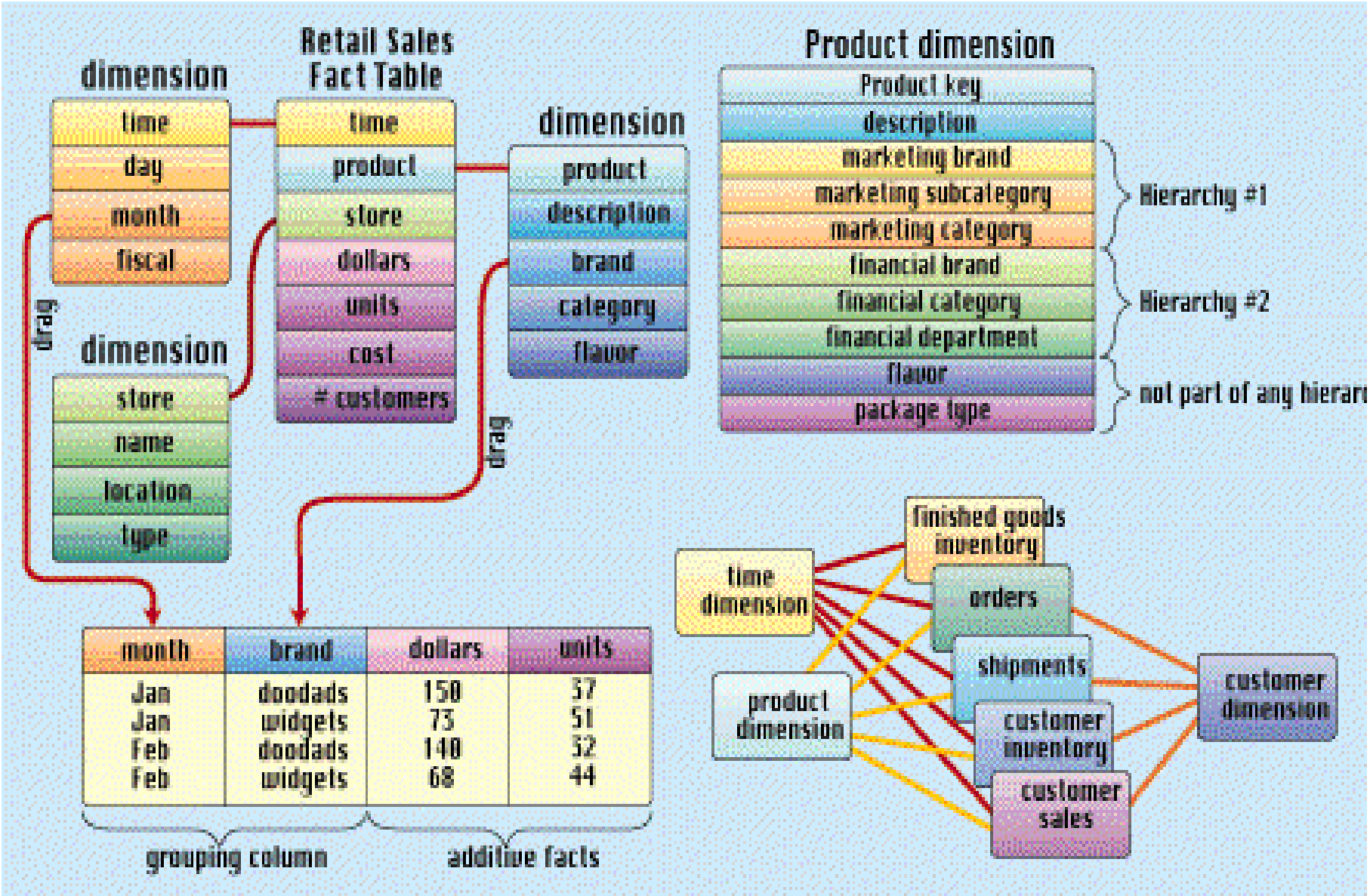
		Brasov	Sibiu	Arad	Iasi	Total
Trim. 1	Mere	-	-	2500	1500	**4000**
	Cirese	-	-	2000	2000	**4000**
	Struguri	1000	3000	-	-	**4000**
	Pepeni	2000	2000	-	-	**4000**
	Total trim 1	**3000**	**5000**	**4500**	**3500**	**16000**
Trim 2	Mere	4000	-	-	-	**4000**
	Cirese	1000	3000	-	-	**4000**
	Struguri	-	-	1500	2500	**4000**
	Pepeni	-	-	2000	2000	**4000**
	Total trim 2	**5000**	**3000**	**3500**	**4500**	**16000**
	Total	**8000**	**8000**	**8000**	**8000**	**32000**

Agregari si granularitate

- **Granularitatea** – reprezinta nivelul de detaliere la care sunt pastrate datele in depozit
- In functie de **cerintele de analiza**, datele se pot pastra la nivel mai detaliat sau mai agregat (depinde de niv. de detaliere a dimensiunilor)
- **Agregarea** datelor- cresterea performantelor DW
- 10 magazine, 100 produse/marca, vanzari saptamanale

Dacă o interogare necesită...	Atunci trebuie parcurse
1 Produs, 1 Magazin, 1 Săptămână	doar 1 înregistrare din schemă
1 Produs, Toate magazinele, 1 Săptămână	10 înregistrări din schemă
1 Marcă, 1 Magazin, 1 Săptămână	100 înregistrări din schemă
1 Marcă, Toate magazinele, 1 An	52.000 înregistrări din schemă

Exemplu


dimension
time
day
month
fiscal
Retail Sales Fact Table
time
product
store
dollars
units
cost
customers
dimension
product
description
brand
category
flavor
Product dimension
Product key
description
marketing brand
marketing subcategory
marketing category
financial brand
financial category
financial department
flavor
package type
Hierarchy #1
Hierarchy #2
not part of any hierar
drag
dimension
store
name
location
type
drag
month
brand
dollars
units
Jan
Jan
Feb
Feb
doodads
widgets
doodads
widgets
150
73
140
68
57
51
52
44
grouping column
additive facts
time dimension
product dimension
finished goods inventory
orders
shipments
customer inventory
customer sales
customer dimension

Depozite de date

Structura depozitului de date (colectie de tabele, vederi, indecsi, sinonime…):

- Schema stea,
- Schema fulg de zapada
- Schema constelatie de fapte

De la relational la multidimensional

Normalizat sau dimensional?

- Exista doua abordari pentru stocarea datelor intr-un depozit de date:
  - **Abordarea normalizata** – Inmon
  - **Abordarea dimensionala** – Kimball
- Aceste doua abordari nu se exclud reciproc, si exista si alte abordari. Abordarile dimensionale accepta normalizarea datelor intr-o anumita masura

Normalizat sau dimensional?

- **Abordarea normalizata** – datele din depozitul de date sunt stocate urmarind regulile de normalizare din bazele de date relationale
- Tabelele sunt grupate dupa **domenii de subiecte** care reflecta categoriile generale de date (de ex: client, produse, angajati etc)
- Principalul **avantaj** al acestei abordari este faptul ca adaugarea de informatii in baza de date este usoara
- **Dezavantaj**: numarul mare de tabele face dificila
  - Combinarea datelor din surse variate
  - Accesarea datelor fara intelegerea semnificatiei surselor de date si structurii depozitului de date

Normalizat sau dimensional?

- **Abordarea dimensionala**: datele sunt impartite in **fapte** (date numerice) si **dimensiuni**

(informatii de referinta care ofera contextul faptelor).

- **Avantaj**: DW este usor de inteles si utilizat. Regasirea informatiei tinde sa fie foarte rapida.
- **Dezavantaje**:
  - Pentru mentinerea integritatii datelor, **incarcarea** in DW din diferite surse operationale este complicate
  - **Modificarea structurii** depozitului de date este dificila, in caz ca compania isi schimba modelul de business

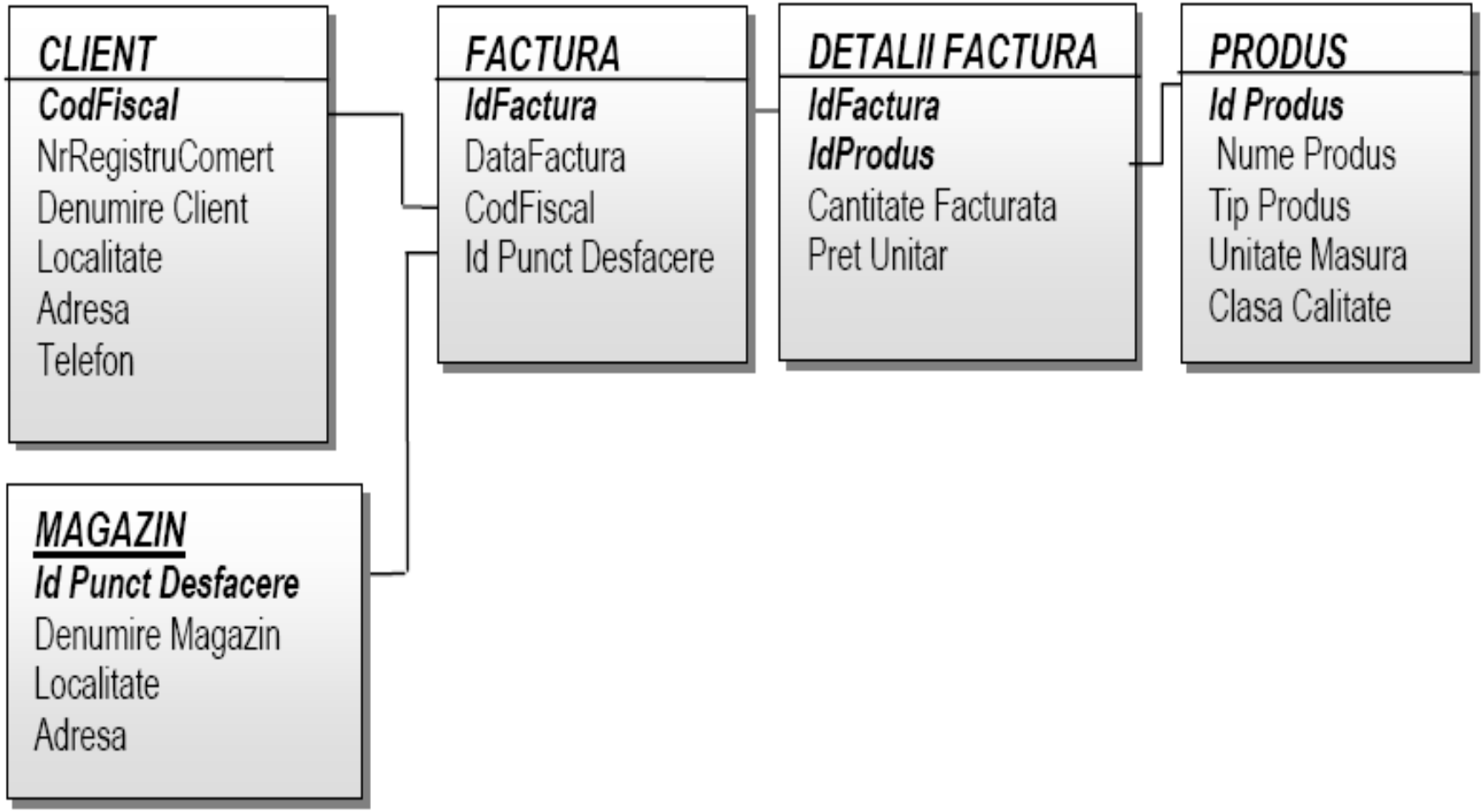
Model relational -Normalizare

- procesul de transformare succesivă a unei BDR în vederea aducerii sale într-o formă standard optimizată
- eliminarea anomaliilor, redundanţelor, dependenţelor nedorite între date
- **Anomalii de actualizare**
  - limitarea posibilităţilor de inserare a datelor
  - pierderi de date la ştergere
  - apariţia de inconsistenţe la modificarea datelor
- **Dependente**
  - ***Dependenţă funcţională*** – A depinde funcţional de un B dintr-o tabelă dacă fiecărei valori a lui A îi corespunde numai o valoare a lui B. B ***depinde funcţional complet*** de un grup de atribute dacă B este dependent funcţional de fiecare atribut din grup.
  - ***Dependentă tranzitivă*** –daca B depinde de A şi C depinde de B atunci C se află în dependenţă tranzitivă faţă de A.
  - ***Dependenţă multivaloare*** – dacă valorii unui atribut A îi corespund două sau mai multe valori ale atributului B

Formele normale

- **Forma normală 1 (FN1)** *dacă* atributele sunt la nivel **atomic** şi au fost eliminate **grupurile de atribute** repetitive
- **Forma normală 2 (FN2)** *dacă* este în FN1 şi nu există **dependenţe funcţionale parţiale** pentru atributele non-cheie
- **Forma normală 3 (FN3)** *dacă* este în FN2 şi nu există **dependenţe funcţionale tranzitive** pentru atributele non-cheie
- **Forma normală 4 (FN4)** *dacă* este în FN3 şi există cel mult o dependenţă funcţională multivaloare pentru atributele non-cheie
- **Forma normală 5 (FN5)** *dacă* este în FN4 şi nu există dependenţe joncţiune pentru atributele non-cheie

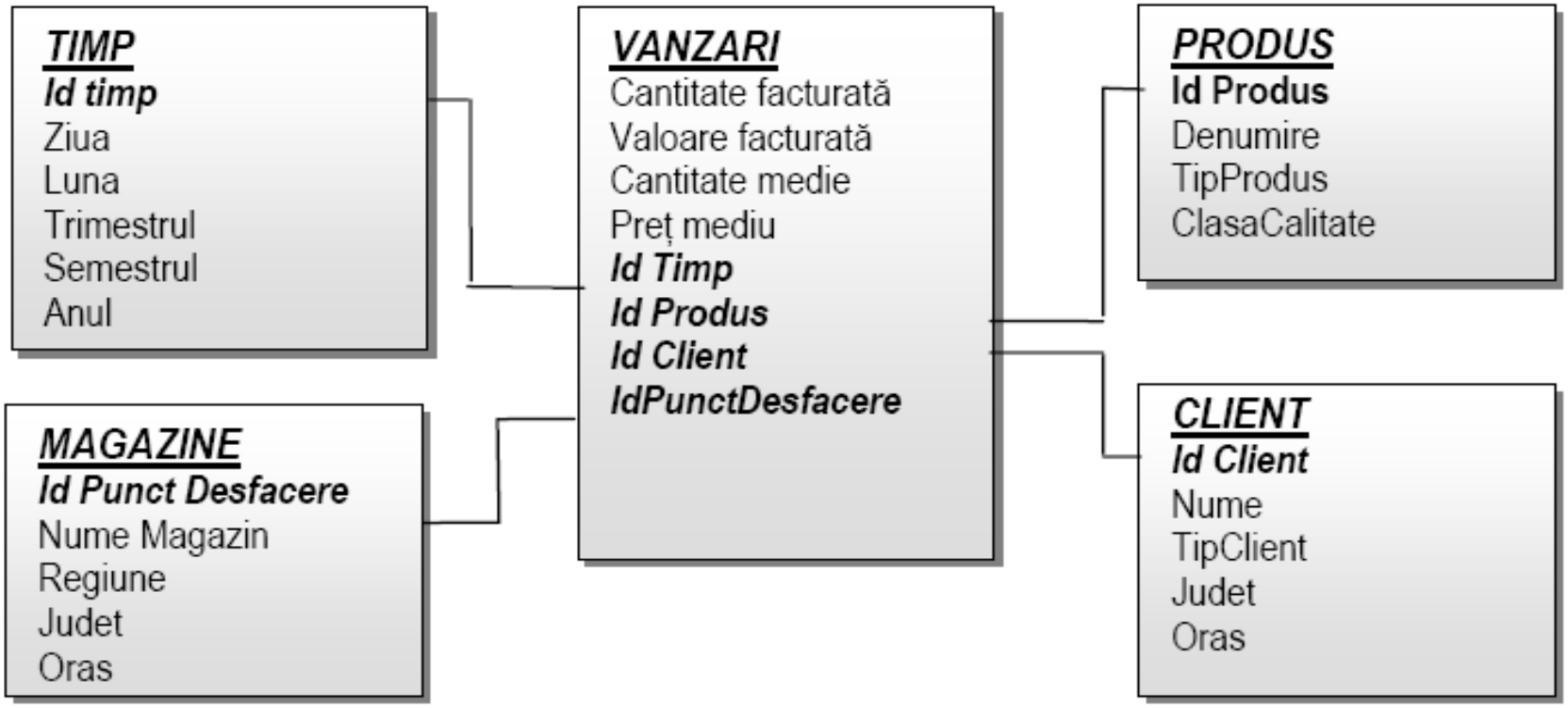
Evidenta facturi – model relational



a.Structura DW – Schema STEA

- cel mai des utilizat model de organizare al depozitelor de date
- **tabela de fapte** cuprinde, fără redundanţe, marea parte a datelor
- tabela de fapte este conectata la tabelele dimensiune pe baza cheilor externe pe care acestea le conţin.
- **star join** = legatura stabilita între un tabel de fapte si tabelele dimensiune
- **star query** = jonctiunea dintre un tabel de fapte si mai multe tabele dimensiune
- **Avantaj**: performante optime pentru interogarile dintr-un depozit de date

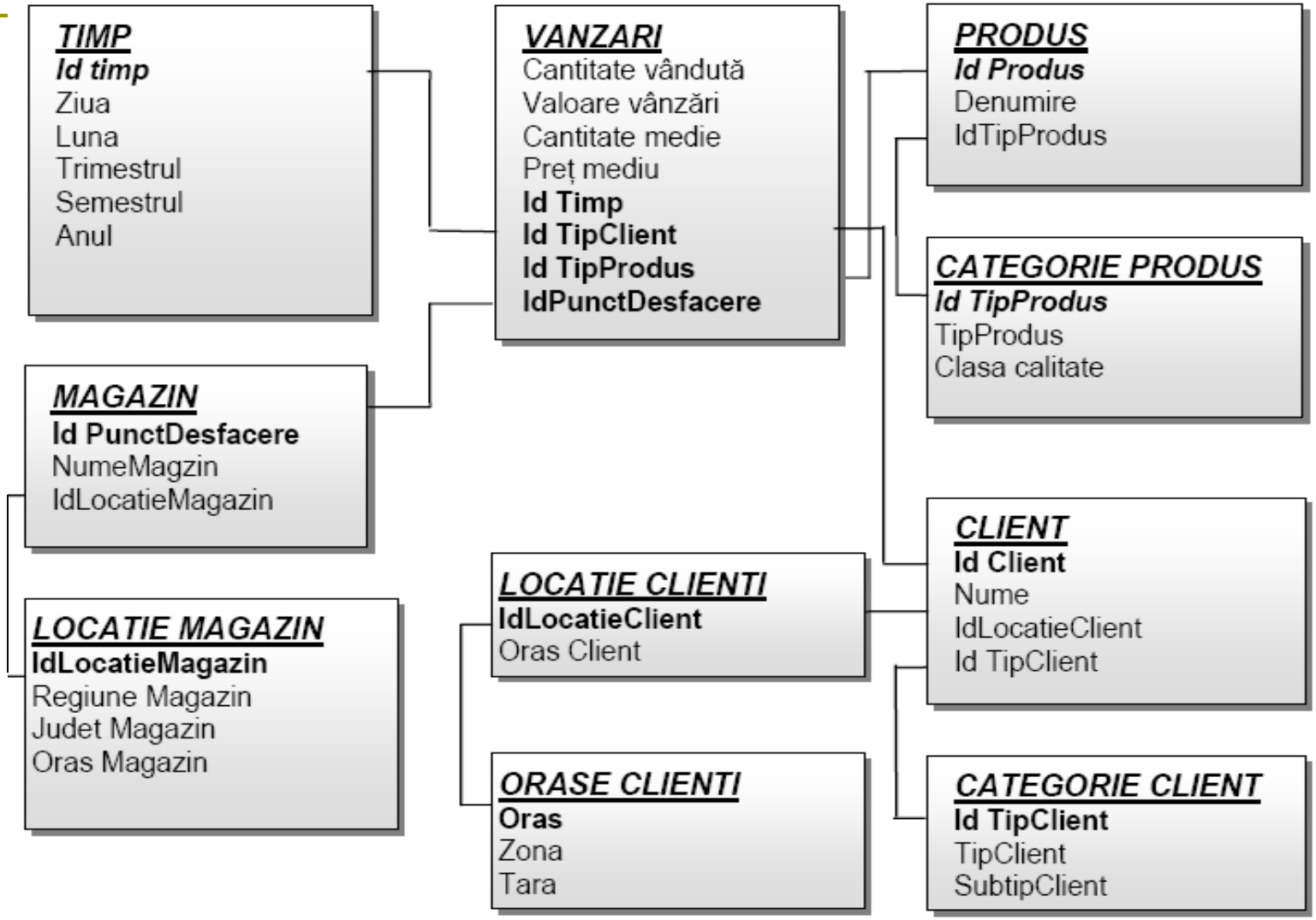
Ex: Schema STEA



b.Structura DW– Schema fulg de zapada

- "**seminormalizat**", avantajele modelului relaţional.
- tabelele dimensiune respecta regulile de normalizare din modelul relaţional =>economie de spaţiu
- nu va conduce la reducerea spaţiului pt tabela de fapte
- **Avantaje**:
  - Redundanta redusa
  - Usor de întretinut
- **Dezavantaje**: la cereri de interogare complexe(join)=> creste timpul de raspuns

Ex: Schema fulg de zapada

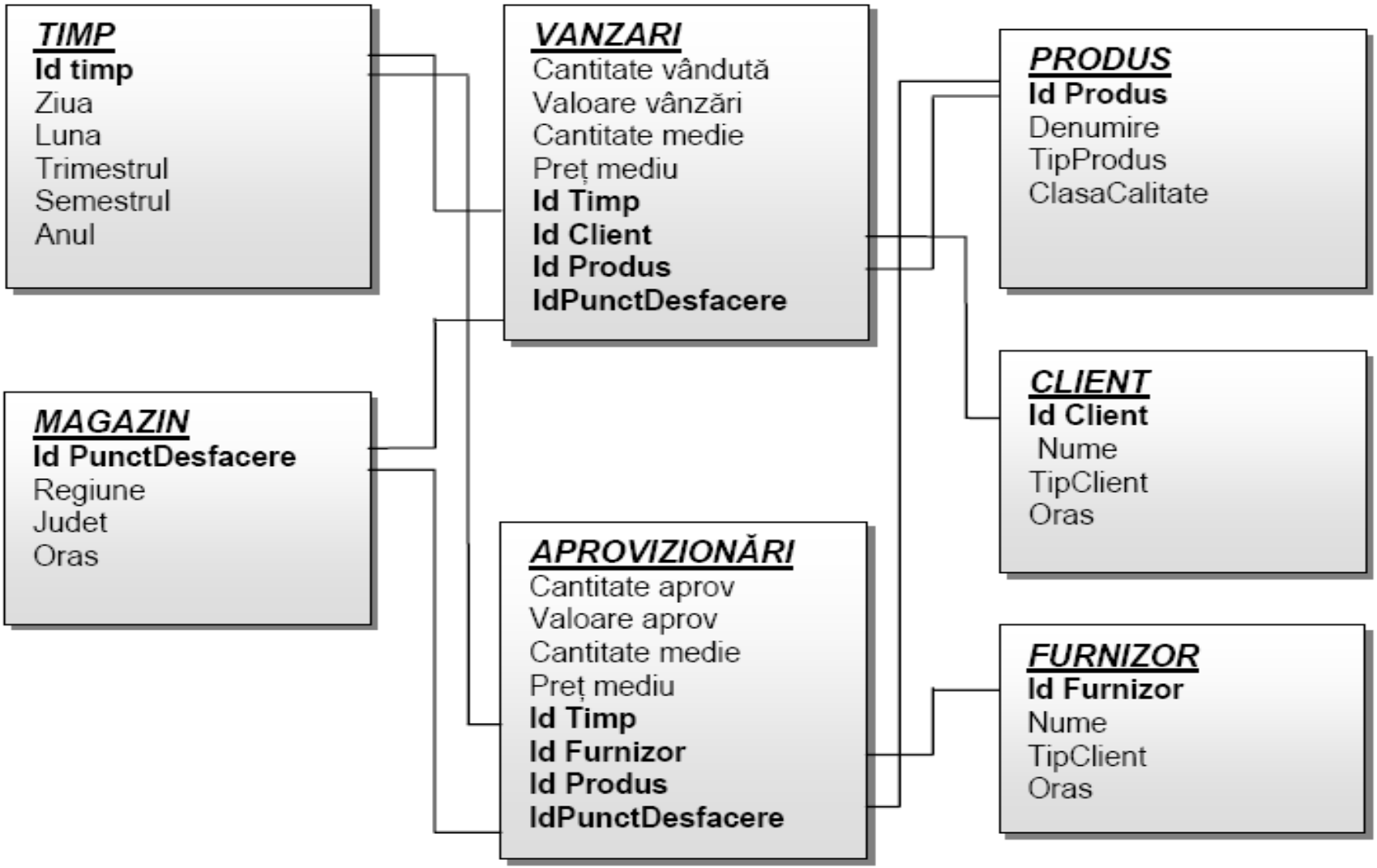


c. Structura DW – Schema constelație de fapte

- Schema galaxie
- mai multe tabele de fapte, conectate ce utilizează aceleaşi tabele-dimensiune

- pe lângă tabela de fapte **Vânzări**, o tabelă suplimentară de fapte **Aprovizionări**, legata de dimensiuni

Ex: Schema constelație de fapte



b. De la relational la multidimensional

- premise diferite, tehnici diferite şi produc BD cu structuri diferite.
- modul de abordare a datelor (utilizator/date):
  - **model multidimensional** - dimensiuni cât mai apropiate de cele naturale şi de perspectiva utilizatorului**.**
  - **model relational** – perspectiva datelor
- model multidimensional:
  - o BD mult mai uşor de consultat şi de interogat la un nivel înalt, sintetic, agregat
  - o BD cu mai putine tabele şi chei de administrat decât modelul relational

Paralela între prelucrarea relațională şi cea analitică

Caracteristici	Modelul relațional	Modelul multidimensional
Organizarea datelor	Tabela	Dimensiuni, tabele de fapte, cub de date
Nivelul datelor	Detaliu	Agregat
Operația tipică	Actualizare	Raportare şi analiză
Nivelul de analiză cerut	Scazut	Ridicat
Volum de date per tranzacție	Redus	Mare
Vârsta datelor	Curente	Istorice, curente, previzionate

Procesul Kimball
1. Selectarea procesului modelat

- **Procesul** este o activitate desfasurata in mod natural de o organizatie
- De obicei, este sprijinit de un sistem de colectare a datelor
- Exemple de **procese de business**:
  - Achizitia de materii prime
  - Gestiunea comenzilor
  - Gestiunea productiei
  - Transportul
  - Gestiunea stocurilor

1. Selectarea procesului modelat (Kimball)

- **NU ESTE** un serviciu sau departament
- Daca modelele dimensionale sunt legate de departamente, vor aparea **duplicari** inevitabile, purtand etichete si terminologie diferita.
- Modelarea mai multor fluxuri de date in modele dimensionale separate vor creste vulnerabilitatea la inconsistenta datelor
- Cea mai buna cale de asigurare a consistentei este **publicarea datelor o singura data** – ceea ce reduce efortul ETL

2. Declararea granularitatii

- Raspuns la intrebarea: **Cum descriu un singur rand din tabela de fapte?**
- **Granularitatea** semnifica nivelul de detaliu asociat masurilor din tabela de fapte
- **Exemple**:
  - Un rand dintr-o reteta primita de la doctor
  - Un rand de pe bonul de casa de la un magazine
  - Un tichet de imbarcare la un zbor
  - Un extras lunar pentru un cont la banca
- Daca la pasii 3-4 se descopera ca granularitatea nu este buna, revenim la 2

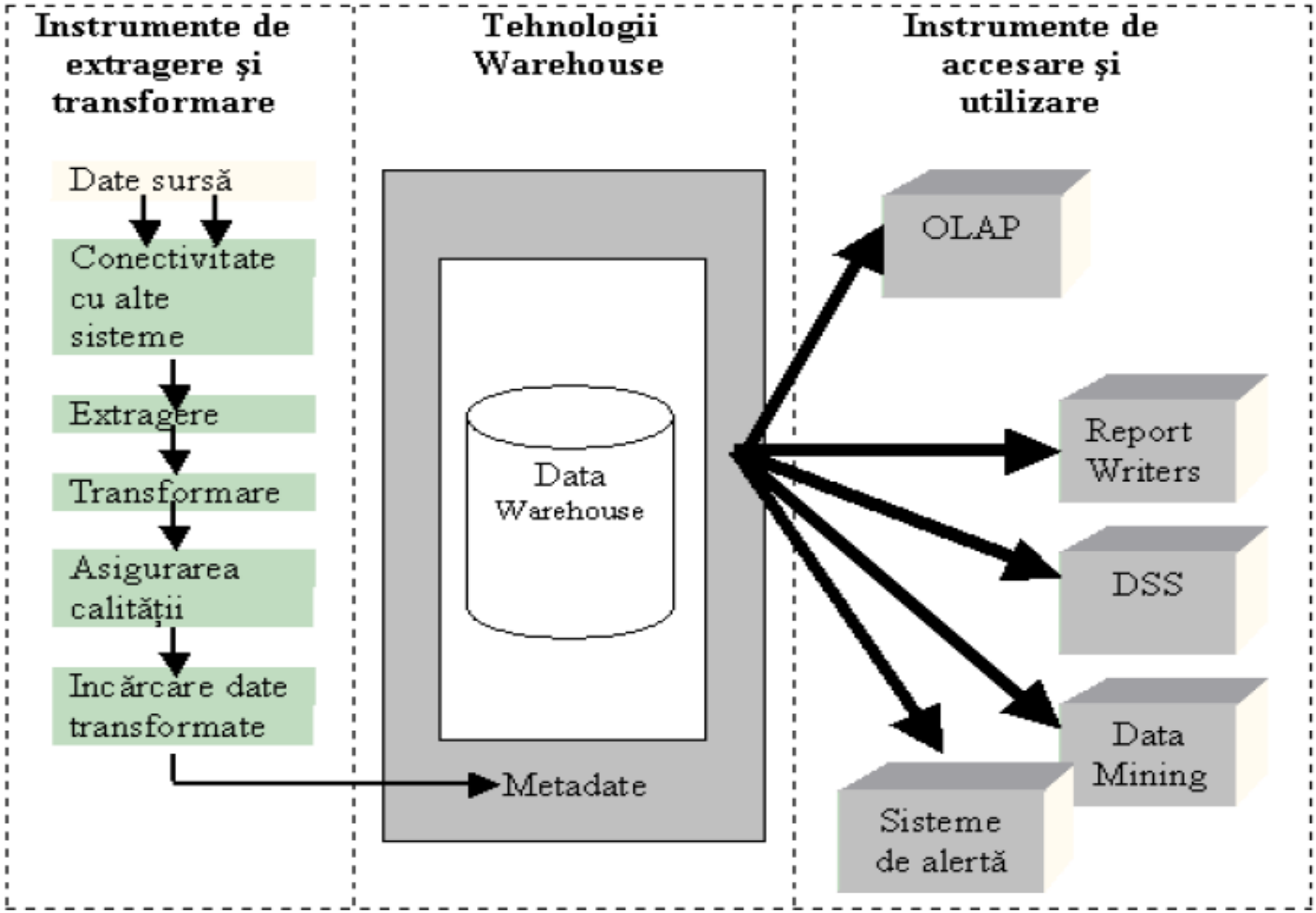
3. Alegerea dimensiunilor

- Daca granularitatea este clara, **dimensiunile** sunt usor de identificat: reprezinta toate descrierile posibile care au valori unice in contextul fiecarei masuratori.
- **Exemple**: data, produs, client, tip tranzactie, stare, etc

4. Identificarea faptelor

- Raspunsul la intrebarea: "**Ce trebuie sa masuram?**"
- Toate faptele candidate trebuie sa fie adevarate la granularitatea definita in pasul 2.
- *Faptele care apatin unui alt nivel de granularitate trebuie separate in alta tabela de fapte*
- De obicei faptele sunt numere aditive: cantitate comandata, valoarea costului in euro etc.

2.Instrumente ETL



ETL

- Acronim pentru
  - E – Extract
  - T –Transform
  - L – Load
- Extragere de date, aplicare reguli de business astfel incat datele sa fie transformate in informatii si stocate in Data Warehouse
- Curatare si standardizare date
- Integrare date interne si externe

ETL(2)

- Simplificare a procesului de copiere a datelor dintr-o sursa in alta
- Datele sunt extrase dintr-o BD tranzactionala, transformate conform **regulilor de business si structurii DW** si incarcate in DW
- Exista posibilitatea incarcarii si din sisteme sursa non-tranzactionale: fisiere, sisteme legacy, sisteme tabelare
- **ETL** trebuie gandit ca si **proces** nu ca si implementare fizica

ETL(3)

- Combinatie complexa de **proces** si **tehnologie** utilizate in procesul de creare a sistemului DW
- Necesita **cunostinte** de analist de business, administrator baza de date si dezvoltator
- Este un **proces recurent**, datele se incarca recurent catre sistemul de DW
- Trebuie sa fie un proces
  - **automatizat**,
  - **bine documentat** si
  - **usor de modificat**

ETL - **Staging Database**

- Operatiile de tip ETL ar trebui efectuate la nivelul unei baze de date relationale, *separata* de sursa de date si de destinatia de date - Data Warehouse
- Creaza o **separatie fizica si logica** intre sistemele sursa si sistemul de Data Warehouse
- **Minimizeaza impactul procesarilor** periodice intense ETL, atat la nivelul sistemelor sursa, cat si la nivelul sistemelor destinatie
- **Nu** permite accesul **utilizatorilor finali**

I. Extragere - Conexiunea cu alte sisteme

- Cel mai dificil aspect este **integrarea sistemelor dispersate**, astfel incat sa fie utilizabile in Data Warehouse
- Datele sunt extrase din sisteme sursa intre care exista diferente la nivel de:
  - SGBD
  - Sisteme de operare
  - Hardware
  - Protocoale de comunicatie
- Exemple:
  - IBM DataJoiner,
  - Oracle Transparent Gateway
  - Sybase Entreprise Connect.

ETL Extragere

- Factori:
  - BD si platforma sistemului sursa;
  - Functionalitatii de extragere si duplicare existente;
  - Intervalele de timp în care sistemele operationale sunt disponibile.
- Metode de baza pentru extragere:
  - **Extragerea in masa** =bulk extraction (intreg depozit)
  - **Replicarea** (doar datele care au fost modificate)
- Curatarea
  - Completarea valorilor lipsa, corectarea erorilor de introducere a datelor, stabilirea unor formate standard, înlocuirea sinonimelor cu identificatori standard
  - Datele recunoscute ca fiind eronate si nu pot fi curatate sunt **respinse**
  - Informatiile culese cu prilejul acestei operatii pot fi folosite pentru **îmbunatatirea calitatii** datelor în timp

Extragere – Tabele de mapare

- Este esential sa existe o **mapare logica** inaintea inceperii implementarii efective
- Maparea trebuie sa furnizeze informatii referitor la extremele transformarii – de obicei reprezentate sub **forma de tabela**

Destinatie			**Sursa**			**Transfor-mare**
Tabela	Coloana	Tip data	Tabela	Coloana	Tip data	

- **Tabelele de mapare** sunt de fapt un **blue-print** pentru dezvoltator
- Tabelele de mapare trebuie sa fie explicative si clare
- Exista o multitudine de tipuri de transformari. De obicei, **exprimate in SQL**

ETL Analiza sistemului sursa

- Este de obicei pasul initial al unui proces ETL
- Poate fi impartit in doua faze:

A1. Faza de descoperire / identificare a datelor

- Criteriul esential de care depinde succesul implementarii este **coerenta si corectitudinea datelor**
- Odata identificata structura rezultatului trebuie analizate si sursele de date

A2. Faza de detectie a eventualelor anomalii

- Esentiala pentru determinarea modalitatii de **tratare a anomaliilor**
- Detectia trebuie urmata de identificare de **proceduri** menite sa minimizeze prezenta si complexitatea anomaliilor

ETL Faza de descoperire / identificare date

- Face parte din atributiile **echipei ETL** – pleaca de la necesarul de date
- Activitatile care trebuiesc efectuate in aceasta faza
  - Identificarea sistemelor sursa
  - Colectarea informatiilor si ***documentarea*** sistemelor sursa
  - Identificarea ***originii datelor*** in cazul existentei surselor multiple si redundantei datelor
  - Intelegerea datelor:
    - Dpdv **tehnic** (gestionare val NULL – atentie la chei externe, gestionare formate diferite),
    - Dpdv **economic**

Schimbari in sursele de date

- Nu sunt importante in momentul incarcarii initiale, dar devin importante pentru incarcarile ulterioare
- Capturarea si urmarirea schimbarilor in sistemele sursa devin o prioritate **pentru incarcarile incrementale** si cad in sarcina echipei ETL
- **Coloane pentru audit**
  - sunt adaugate la fiecare tabela pentru a stoca **data si ora** la care o inregistrare a fost inserata sau modificata
  - trebuie analizate si testate atent pentru a vedea daca sunt o sursa de incredere pentru a indica schimbarea datelor

Determinarea datelor modificate

- Procesul de eliminare pastreaza o singura copie a fiecarei extrageri anterioare in **staging area**
- In timpul incarcarii urmatoare, procesul preia tabelele sursa in intregime in staging area si face **o comparatie** cu datele pastrate de la ultima incarcare
- Doar **diferentele** sunt trimise in DW.
- Nu este cea mai eficienta tehnica, dar este cea mai de incredere pentru capturarea schimbarilor datelor

Determinarea datelor modificate – incarcari initiale si incrementale

- Se creeaza doua tabele:
  - O tabela cu incarcarea **anterioara** si
  - O tabela cu incarcarea **curenta**
- Procesul de **incarcare initiala in masa** incarca date in tabela de incarcare curenta. Nu se aplica detectarea schimbarilor, ci datele sunt transformate si incarcate direct in tabelele tinta.
- Cand procesul se termina, el sterge tabela cu incarcarea anterioara si redenumeste tabela de incarcare curenta ca tabela de incarcare anterioara
- **La urmatoarea executie** a procesului, tabela de incarcare curenta este populata
- Se selecteaza tabela curenta de incarcare **MINUS** tabela cu incarcarea anterioara; se transforma si se incarca in DW doar setul de date rezultat

II. Transformare

- Functii oferite:
  - Partitionarea si consolidarea câmpurilor
  - Standardizarea
  - Deduplicarea.

Sistem sursa	**Tipul transformarii**	**Depozit de date**
Câmpul Adresa Str. Unirii Nr. 123, Municipiul Iasi, 6600, România	Partitionare câmpuri	**Nr. Str.**: 123 **Strada**: Unirii **Localitate**: Iasi **Tip localitate**: Municipiu **Cod Postal**: 6600 **Tara**: România
Sistem A, **Functie**: Manager general **Sistem B,** **Functie**: Director general	Consolidare câmpuri	**Functie**: Manager general sau Director general
Data comenzii: 21 Nov. 2002 **Data comenzii**: 01-09-02	Standardizare	**Data comenzii:** 21 Noiembrie 2002 **Data comenzii**: 01 Septembrie 2002
Sistem A, **Nume angajat:** Popescu I. Vasile **Sistem B,** **Nume angajat:** Popescu Vasile	Deduplicare	**Nume angajat:** Popescu I. Vasile

ETL - Transformare

- Este pasul principal in care se aplica **seturi de reguli de business** identificate
- Este pasul principal in care este **adaugata valoare** in procesul de ETL
- Este singurul pas in care **datele sunt efectiv modificate** in acest proces
- Este implementat **la nivelul Staging Database**
- Aici trebuie implementate elemente de **validare a calitatii datelor**
- Datele trebuie sa fie
  - Corecte
  - Cu grad de ambiguitate minim
  - Consistente
  - Complete

ETL - Transformare

- **Analiza calitativa a datelor** – in minim 2 momente in cadrul ETL (extractie si transformare)
  - Detectie anomalii – teste pe esantioane de date
  - **Validari** la nivel de **camp**
    - Valoare NULL
    - Valori numerice care ies din tiparele standard permise
    - Valori care nu se incadreaza in plaja de valori admise
    - Valori care nu urmaresc template-urile utilizate
  - **Validari structurale** la nivel de **tabela**
    - Cheile tabelelor sunt definite corect
    - Restrictia de integritate este satisfacuta
  - **Alte validari**
    - **Validari** ale **logicii de business**

Motive pentru date “murdare”

- Prezenta valorilor “dummy”
- Absenta datelor
- Campuri utilizate in mai multe scopuri
- Date criptate
- Date contradictorii
- Utilizarea gresita a anumitor campuri in sistemele sursa (vezi campuri de tip adresa)
- Violarea regulilor de business
- Reutilizarea cheilor primare
- Utilizarea identificatorilor non-unici
- Probleme la integrarea datelor

Curatarea datelor

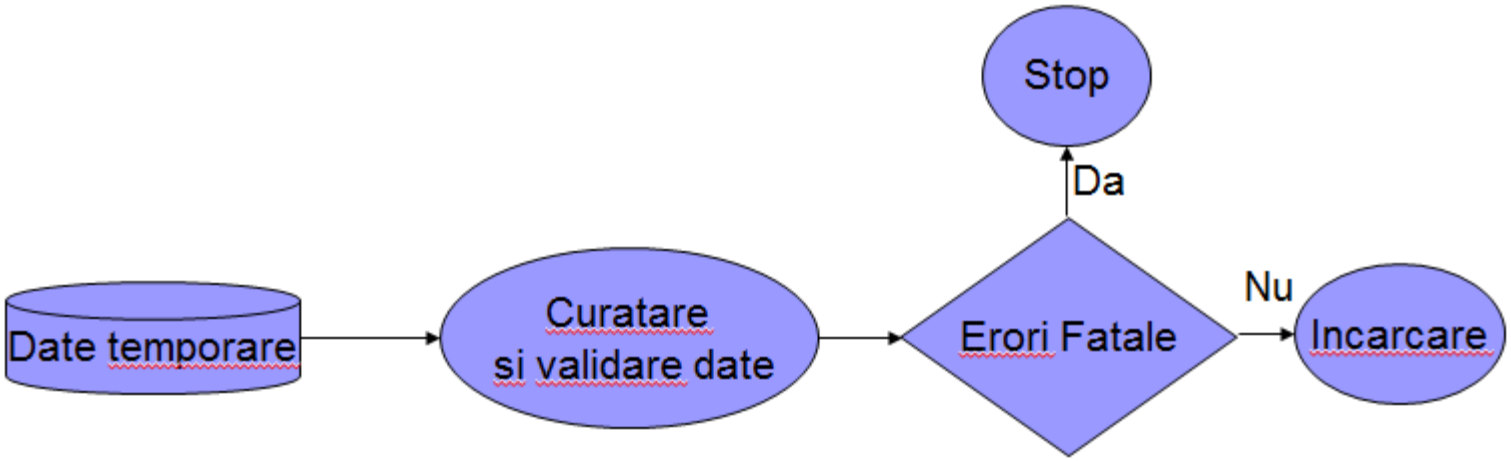
- **Partitionare/ Parsing**
  - Identificarea **campurilor individuale** in cadrul surselor de date si **izolarea** acestora in cadrul destinatiei. Exemplu: campuri de tip adresa
- **Corectie**
  - Faza in care eventualele **anomalii sunt eliminate** prin utilizarea algoritmilor complecsi sau a altor surse de date. Exemplu, determinare cod postal
- **Standardizare**
  - Faza in care datele sunt stocate intr-o **forma unica**, preferata, aplicand o multitudine de reguli

Curatarea datelor

- **Potrivire/ deduplicare**
  - Pasul de cautare si **imperechere** a inregistrarilor care vizeaza o aceeasi entitate fizica in scopul **eliminarii duplicatelor**. Exemplu: cautare, identificare si imperechere inregistrari care se refera la o aceeasi persoana insa al carui nume este stocat diferit
- **Consolidare**
  - **Eliminarea** efectiva a **duplicatelor** identificate in urma aplicarii regulilor detaliate anterior

ETL

- Transformare



Instrumentele pentru asigurarea calității datelor

- Asista la **localizarea si corectarea erorilor** in sistemele sursa sau DW
  - In sistemele sursa - preferabil
  - In depozitul de date - inconsistente
- **Pana la 15%** din datele extrase sunt inconsistente sau incorecte
- Exemple
  - Data Quality Workbench (DataFlux);
  - Content Tracker (Pine Cone Systems);
  - Quality Manager (Prism)
  - Integrity Data Reengineering (Vality Technology)

III.Incarcarea datelor

- Ajuta la incarcarea datelor transformate in depozitul de date
- **Preformatarea** datelor în formatul fizic intern cerut de SGBD-ul tinta
- Trebuie sa asigure **integritatea** si **consistenta** datelor preluate din sistemele sursa
- Este cel mai **mare consumator de timp**
  - Datele sunt stocate in *tabele denormalizate*
  - **Indecsii** pot încetini substantial procesul de încarcare – se renunta la ei înainte de încarcare si apoi se recreaza
  - Permisa doar in anumite intervale orare

ETL Incarcarea datelor modificate

- **Incarcari initiale, complete**
  - Utile in cazul in care volumul de date nu este considerabil
  - Se extrag din sistemul sursa toate inregistrarile prezente in momentul extractiei
- **Incarcari incrementale**
  - Utile in cazul volumelor mari de date
  - Se extrag din sistemul sursa doar inregistrarile actualizate (nou create, modificate, sterse) de la ultima incarcare si pana la momentul extractiei
- Instrumentele I,II,III sunt de obicei incorporate în cadrul unui singur instrument, **ETL Tools**
- **Exemple:** vezi figura

Informatica PowerCenter
IBM InfoSphere Datastage
SAP Business Objects Data Integrator
Oracle Data Integrator
SAS Data Integration Studio

Figure 1. Magic Quadrant for Data Integration Tools