

EXEMPLU PROBLEME

1) Exemplu Apriori:

Se da urmatorul set de tranzactii extrase din vanzarile unui magazin alimentar:

$D = \{(1, [\text{oua}, \text{paine}, \text{lapte}]), (2, [\text{cartofi}, \text{paine}, \text{bere}]), (3, [\text{oua}, \text{cartofi}, \text{paine}, \text{bere}]), (4, [\text{cartofi}, \text{bere}])\}$

Semnificatia este urmatoarea : D contine 4 tranzactii (de la 1 la 4), fiecare tranzactie contine o serie (continuta intre [si]) de alimente cumparate (oua, cartofi, paine, lapte, bere).

Folosind algoritmul Apriori pentru gasirea regulilor de asociere sa se gaseasca :

a. (1p) Setul/seturile de elemente de lungime maxima ('large itemsets') care sa aiba gradul de suport minim de 50% ('support').

b (0.5p) Din aceste seturi de elemente de lungime maxima ('large itemsets'), sa se genereze regulile de asociere care au gradul de incredere minima de 75% ('confidence').

2) Exemplu map-reduce:

Se da urmatorul set de tranzactii extrase din vanzarile unui magazin:

$D = \{[\text{cartofi}, \text{morcovi}, \text{rosii}], [\text{ardei}, \text{morcovi}, \text{ceapa}], [\text{cartofi}, \text{ardei}, \text{morcovi}, \text{ceapa}], [\text{ardei}, \text{ceapa}]\}$. Folosind algoritmul Map-Reduce, contorizati fiecare tip de articole vandut din setul de tranzactii dat.

3) Problema DW - ETL - Preprocesare - OLAP:

În baza de date relatională există tabela de tranzacții zilnice *DailyOrders* = {*storeId*, *date*, *orderId*, *amount*}. Tabela *Store* conține un câmp *location* (județul). La sfârșitul lunii, datele trebuie consolidate într-o tabelă partitionată (după 4 regiuni geografice: NordEst, Sud, Centru, Vest,), din depozitul de date cu structura *Sales* = {*region*, *year*, *month*, *amount*}.

- a) (1pct) Scrieți clauzele de partitionare pentru tabela *Sales* din depozitul de date;
- b) (1pct) Scrieți instrucțiunile SQL pentru operațiile ETL de actualizarea lunară a tabelului *Sales* (locatiile se centralizează pe regiuni, conform tabelului de mapare *Store*)
- c) (1pct) Scrieți instrucțiunile SQL pentru operația de tip “rollup” pentru agregarea vânzărilor la nivel de trimestru. Scrieți instrucțiunile SQL pentru operația de tip “dice” pentru aflarea valorilor vânzărilor pe ultimul trimestru al anului trecut pentru regiunea “NordEst”.

4) Problema NoSQL: MongoDB - OrientDB

Modelați pe OrientDB (modelul graf) problema gestiunii studenților (*grupa*, *nume*, *prenume*), a disciplinelor de curs (*nume*, *cod*, *profesori*) și a evaluării (*disciplina*, *student*, *nota*) de către profesori (*nume*, *prenume*).

- a. (2pct) Reprezentați grafic modelul graf. Scrieți clasele necesare modelului, precum și structura a 6 vertexi, respectiv edge-urile corespunzătoare ce pot fi inserați/inserate în baza de date.
- b. (1pct) Afișați rezultatele studenților la disciplina ECBD ordonați crescător grupe și în ordinea descrescătoare ale notelor obținute.