

# C13

## SQL si JDBC

- Baze de date, concepte, modele de date
- Structured Query Language:
  - DDL, DML
  - exemple
- Java DataBase Connectivity:
  - motivare
  - integrare
  - arhitectura
  - utilizare

## Obiective

După parcurgerea acestui curs ar trebuie sa puteți:

- înțelege conceptele vehiculate in baze de date si cele legate de modelul de date relational;
- intelege bazele si limitele limbajului SQL si cum se utilizeaza in manipularea datelor;
- intelege rolul si necesitatea utilizarii limbajelor de nivel inalt in scrierea aplicatiilor ce opereaza cu baze de date;
- intelege si utiliza clasele cu interfetele JDBC pentru accesul la bazele de date din Java.

# Concepte legate de bazele de date relational si modele de date

- *Relație (relation)* - în **matematică**, relația  $R$ , este o submulțime a produsului cartezian a două mulțimi  $(A, B)$ ,  $R : A \times B$ . Dacă  $(a, b)$  este un element din  $R$  notația  $a R b$  are semnificația de "a este în relația  $R$  cu b" (o notație echivalentă folosită, mai des, în teoria bazelor de date relaționale este  $R(a, b)$ ). O relație poate fi *reflexivă* ( $a R a$ ), *simetrică* ( $a R b \Rightarrow b R a$ ), *transitivă* ( $a R b \ \& \ b R c \Rightarrow a R c$ ), *antisimetrică* ( $a R b \ \& \ b R a \Rightarrow a = b$ ) sau *totală* ( $a R b$  sau  $b R a$ ). În **bazele de date**, relația, este *un tabel* dintr-o bază de date relațională.
- *Bază de date (database)* - una sau mai multe mulțimi de date persistente asociate cu o aplicație pentru actualizarea și interogarea datelor;
- *Persistentă* - proprietate a unei aplicații de a crea obiecte și date care continuă să existe și să-și păstreze valorile între două execuții ale aplicației.
- *Model de date (data model)* - formalism matematic cu notații pentru descrierea structurilor de date și a mulțimii operatorilor folosiți pentru manipularea și validarea datelor;
- *Model de date relațional (relational model)* - un model de date introdus de E. F. Codd în 1970. În acest model datele sunt organizate în tabele. Mulțimea numelor coloanelor se numește schema tabelului. Datele pot fi manipulate folosind algebra relațională. Limbajul SQL este implementarea unei astfel de algebre.
- *SQL (Structured Query Language)* - un limbaj neprocedural, de interfață, între utilizator și bazele de date relaționale dezvoltat de IBM în anii '70. Diferența esențială între SQL și alte limbaje este aceea că în SQL instrucțiunile specifică ce operații se vor face cu datele și nu modul în care se realizează acestea.
- *Bază de date relațională (relational database, relational database management system)* - o bază de date ce utilizează modelul de date dezvoltat de E. F. Codd. O bază de date relațională permite definirea structurilor de date, operațiile de stocare, de extragere a datelor și a constrângerilor de integritate. Într-o bază de date relațională (BDR), datele și toate relațiile între date sunt organizate în tabele. Un tabel este o colecție de înregistrări (rânduri) unde fiecare înregistrare a tabelului conține aceleași câmpuri. Anumite câmpuri pot fi desemnate chei. În acest caz viteza de căutare a unor valori specifice aceluși câmp va crește datorită folosirii de indexuri. Între înregistrările unor tabele diferite se pot realiza asocieri pe baza valorilor comune ale unui anumit câmp particular din fiecare tabel.
- *Cheie primară* - numărul minim de câmpuri care au fost alese pentru a identifica unic fiecare rând dintr-un tabel.
- *Constrângere de integritate* - o regulă care trebuie să rămână adevărată pentru ca baza de date să-și păstreze integritatea. De exemplu, într-o bază de date genealogică fiecare individ

trebuie să fie copilul unor părinți. De asemenea, individul nu poate avea decât doi părinți naturali.

- *Index* - o secvență de perechi (cheie, pointer) în care fiecare pointer poartă la o înregistrare din baza de date care conține valoarea cheii într-un câmp particular. Indexul este sortat pe baza valorilor de chei și permite căutarea rapidă a unei valori particulare de cheie.
- *Tabel* - un obiect al bazei de date care stochează datele. Este format din mai multe coloane. Fiecare coloană are asociat un tip de dată. Tipul coloanei se folosește pentru manipularea corectă a conținutului.
- *Normalizare* - o relație dintr-o bază de date relațională se zice că este într-o formă normală dacă satisface un anumit grup de constrângeri. Lucrarea inițială a lui Codd definea trei forme normale.
- *Tuplă* - un obiect de date ce conține una sau mai multe componente. Componentele unei tuple pot fi de tipuri diferite. Câteva exemple de tuple sunt: (1,2), ("Tuple", Da), (a, (x, y), c).
- *Tranzacție* - o unitate logică de interacțiune cu o BDR. Aceasta trebuie tratată într-o manieră coerentă și sigură, independent de alte tranzacții. Pentru o tranzacție se garantează, fie realizarea ei completă, fie nerealizarea ei. Dacă o eroare conduce la realizarea parțială a tranzacției, tranzacția poate fi "derulată înapoi" pentru a opri lăsarea bazei de date într-o stare de inconsistență.
- *Interogare (query)* - acțiune de extragere a unor informații din baza de date.

## Conceptul de dată

Pentru cei care au deja experiență în domeniul calculatoarelor, noțiunea de dată este asociată cu numere, șiruri de caractere, imagini, sunete ce pot fi stocate și prelucrate pe calculator, eventual transmise mai departe prin rețea. Privind însă conceptul de dată general, data, reprezintă o reflectare, parțială, a realității. Deseori, cuvântul de dată apare în contextul conceptului de model. Necesitatea modelului apare datorită complexității lumii pe care încercăm să o surprindem. Abundența detaliilor ne depășește. Din acest motiv, realitatea va fi simplificată de om, prin crearea unor modele. Modelul se obține prin eliminarea detaliilor și păstrarea unor caracteristici minimale din realitate, fără a pierde însă veridicitatea modelului (care dintre caracteristici sunt păstrate, deoarece se consideră a fi semnificative, respectiv de ce și care anume se elimină asta este o altă carte). Datele reprezintă o măsură cantitativă a procesului de evoluție a caracteristicilor considerate semnificative pentru model (aceste caracteristici descriu modelul). Generalizând, datele reprezintă o colecție de observații pe marginea unor evenimente pe care le considerăm importante.

Datele în sine nu au nici o semnificație (ele nu au un înțeles). Ele pot însă deveni informație în urma asocierii directe acestora cu anumite caracteristici. Datele vor fi astfel încărcate cu semnificație ca urmare a interpretării lor într-un context particular. Această interpretare se poate face de către om sau, automat, de către un sistem de calcul. Dacă considerăm numărul întreg "1100", el este dată dar nu este informație. În contextul "Salariul meu este de 1100 RON", data "1100" are deja un înțeles particular (este valoarea remunerării mele), devenind informație. În contextul "Salariul meu de 1100 RON este acceptabil.", data "1100" permite trecerea la cunoaștere. Deoarece eu cunosc salariile unei mari majorități, sunt în stare să calculez un salar mediu și să mă plasez undeva, în raport cu acesta. Procesul de plasare la nivelul de "acceptabil" are la bază un grup de prelucrări iar rezultatul de "acceptabil" are la bază un proces deductiv. Foarte pe scurt, datele reprezintă materialul primar,

informația reprezintă date + înțeles, iar cunoașterea este informație + prelucrare.

Scopul colectării datelor este obținerea de informații sau de cunoaștere ca urmare a unui proces de deducție bazat pe experiență sau pe reguli ce poate fi implementate la nivelul unui sistem de calcul. În bazele de date, data, va fi organizată sub forma unui singur articol de informație (o singură valoare). Deoarece în procesul de colectare a datelor pot să apară erori, termenul de integritate a datelor definește mecanismele de prevenire a erorilor.

### **Modelul de date relațional**

Se numește model de date, un formalism matematic cu notații pentru descrierea structurilor de date și a mulțimii operatorilor folosiți pentru manipularea și validarea datelor. Modelul de date relațional este introdus de E.F. Codd în 1970. În acest model datele sunt structurate (organizate) în relații care au nume. Mulțimea numelor coloanelor se numește schema relației. Structura de date logică numită relație este bidimensională, fiind cunoscută sub numele de tabel, în baza de date fizică. Cele două dimensiuni ale structurii sunt denumite rânduri și coloane. Datele modelului se pot modifica folosind operatorii algebrei relaționale (limbajul SQL este implementarea unei astfel de algebre), ei asigurând selectarea, inserarea, modificarea și ștergerea datelor din tabele, iar restricțiile de integritate asigură păstrarea consistenței datelor.

### **Conceptul de redundanță**

Redundanța datelor apare atunci când o anumită dată este stocată în mai multe tabele ale bazei de date. Acest surplus de date necesită o atenție specială atunci când se aduc modificări asupra conținutului valorilor stocate în tabele. Dacă valoarea unei date redundante este modificată, atunci va trebui să ne asigurăm că toate copiile respectivei date au fost aduse la nouă valoare. În această situație se zice că valoarea respectivă este consistentă în baza de date. Dacă nu toate copiile au fost actualizate la noua valoare se zice că data este inconsistentă. Se zice că o bază de date este în stare consistentă dacă toate datele acesteia sunt consistente, altfel este în stare inconsistentă.

### **Terminologii alternative ale modelului relațional**

Termenii folosiți în modelul de date relaționali sunt, deseori, producători de confuzie. Deoarece modelul are ca sursă algebra relațională, matematicienii preferă să folosească anumite denumiri. Proiectanții bazelor de date au și ei termeni specifici, dar echivalenți cu cei din algebra. În final, administratorii de sistem folosesc și ei propriul lor jargon atunci când realizează operații de întreținere a bazelor de date. Tabelul care urmează își propune să prezinte termenii alternativi folosiți de cele trei categorii de persoane care folosesc acest model.

<b>Algebra relațională</b>	<b>Proiectant BDR</b>	<b>Administrator de sistem</b>
Relație	Tabel	Fișier
Tuplă	Rând	Înregistrare
Atribut	Coloană	Câmp

**Atributele** reprezintă date elementare care sunt nume ale coloanelor unei relații. De exemplu, relația `chirie(IDClient, NumeClient)` conține atributele `IDClient, NumeClient`. Valorile actuale pentru

aceste atribute ale relației se stochează în **tuple** sau rânduri ale tabelului. Atributele se grupează pe baza dependenței de valorile din cheia primară. O **cheie primară** (CP) este un atribut, sau un grup de atribute, care indentifică unic un rând din tabel. Un tabel poate avea o singură cheie primară. Deoarece valorile din CP se folosesc pentru identificare, ele nu pot fi vide (nule). Asocierea între două relații se face, tipic, printr-un atribut comun celor două relații. Atributul comun este, de obicei, CP într-un tabel și **cheie străină** (CS) în celălalt. Regulile de **integritate referențială** dictează valorile CS dintr-o relație în raport cu valorile CP din cealaltă relație.

Proiectarea unei baze de date relaționale se face pe baza **regulilor de normalizare** care dictează apartenența atributelor la relații. Aceste reguli **nu fac obiectul acestui curs însă** folosirea modelului de date relațional normalizat asigură minimizarea redundanței datelor împreună cu protecția contra anomaliilor de inserare, ștergere și actualizare care apar ca urmare a definerii unor relații incorecte.

# Elemente de SQL

SQL este un limbaj neprocedural (este implementarea unei algebre relaționale) dezvoltat în baza răspândirii modelului relațional. În ultimii ani a devenit limbajul standard pentru BDR. Obiectivele limbajului sunt:

- crearea bazelor de date și a relațiilor;
- realizarea operațiilor de întreținere a BDR;
- realizarea interogărilor simple și complexe.

Ca urmare a standardizării ISO limbajul SQL are două componente majore:

- limbajul pentru definiția datelor (*Data Definition Language - DDL*): pentru definirea structurii bazei de date și controlul accesului la date;
- limbajul pentru manipularea datelor (*Data Manipulation Language - DML*): pentru extragerea și actualizarea datelor.

O *instrucțiune SQL* este formată din *cuvinte rezervate* și *cuvinte definite de utilizator*. Cuvintele rezervate (numite și cuvinte cheie) sunt partea fixată a limbajului deoarece au o semnificație predefinită. Ele trebuie scrise fără a fi despărțite pe mai multe linii. Cuvintele definite de utilizator sunt "inventate" de el și reprezintă nume ale obiectelor bazei de date cum sunt tabelele, coloanele etc. Cuvintele unei instrucțiuni SQL se scriu conform mulțimii regulilor de sintaxă specifice limbajului SQL. Multe dintre dialectele SQL folosesc caracterul "punct și virgulă" ";" pe post de terminator de instrucțiune SQL, însă conform standardului, acestea sunt opționale.

Pentru exemplele care vor fi dezvoltate în continuare este suficientă prezentarea a numai patru dintre instrucțiunile DML din SQL:

1. `SELECT` - pentru interogarea bazei de date;
2. `INSERT` - pentru inserarea datelor într-un tabel;
3. `UPDATE` - pentru actualizarea datelor unui tabel;
4. `DELETE` - pentru ștergerea datelor unui tabel.

## Instrucțiunea `SELECT`

Permite extragerea și afișarea datelor din unul sau mai multe tabele.

### *SELECT cu un singur tabel*

Pentru extragerea tuturor numelor de proiecte stocate în câmpul `NumePr` din tabelul **Proiecte** scriem:

```
SELECT NumePr FROM Proiecte;
```

Pentru extragerea numelor de proiecte și a sumei plătite pe proiect, stocată în câmpul `Suma`, din același tabel se scrie:

```
SELECT NumePr, Suma FROM Proiecte;
```

Deoarece multe interogări SQL folosesc toate câmpurile unui tabel. Folosirea unui asterisc "\*" este o formă prescurtată de scriere pentru "toate câmpurile".

```
SELECT * FROM Proiecte;
```

Atunci când dorim să impunem condiții cu privire la rândurile extrase se va folosi clauza `WHERE`. De exemplu, pentru extragerea proiectelor din tabelul **Proiecte** care au o suma mai mare de 500000 în câmpul `Suma` se va scrie:

```
SELECT * FROM Proiecte WHERE Suma > 500000;
```

Dacă dorim ca rândurile sortate să fie în ordinea crescătoare alfabetică a proiectelor vom folosi clauza `ORDER BY` astfel

```
SELECT * FROM Proiecte ORDER BY NumePr;
```

Pentru sortarea în ordinea descrescătoare a numelor de proiecte scriem:

```
SELECT * FROM Proiecte ORDER BY NumePr DESC;
```

### *SELECT cu mai multe tabele*

Dacă datele de extras sunt stocate, ca urmare a normalizării, în mai multe tabele, una dintre metodele pentru conectarea tabelelor pe baza unui câmp comun folosește instrucțiunea `JOIN`. Să presupunem că dorim să extragem toate proiectele și tipurile membrilor pe baza câmpurilor `IDPr` din **Proiecte** și `IDPr` din **Membri**. `IDPr` este CP în **Proiecte** și CS în **Membri**. Între proiecte și membri este o *relație unu-la-multi*, adică unui rând din tabelul **Proiecte** pot să-i corespundă mai multe rânduri ale tabelului **Membri**. Instrucțiunea `INNER JOIN` face ca din cele două tabele să fie extrase doar rândurile care au valori comune în câmpurile de conectare.

```
SELECT NumePr.Proiecte, Membri.Calitate
FROM Proiecte INNER JOIN Membri ON Proiecte.IDPr = Membri.IDPr;
```

### **Instrucțiunea INSERT**

Instrucțiunea `INSERT` permite adăugarea unui rând într-un tabel. Una dintre formele ei este:

```
INSERT INTO NumeTable (lista de câmpuri)
VALUES (lista de valori)
```

Dacă dorim să inserăm un nou proiect în tabelul **Proiecte** scriem:

```
INSERT INTO Proiecte (IDPr, NumeProiect, Suma1)
VALUES ('PR4', Proiect nou', 100000);
```

### **Instrucțiunea UPDATE**

Instrucțiunea `UPDATE` permite modificarea conținutului unui rând existent dintr-un tabel. Forma ei

este:

```
UPDATE NumeTabel  
SET NumeColoana1=Valoare1 [,NumeColoana2=Valoare2 ...]  
[WHERE Conditie]
```

Pentru a crește cu 20% valoarea din câmpul `Suma` ale tuturor proiectelor din tabelul **Proiecte** scriem:

```
UPDATE Proiecte SET Suma=Suma*1.20;
```

Pentru a crește valoarea din câmpul `Suma` cu 20% a proiectului cu numele "Proiect1" din tabelul **Proiecte** scriem:

```
UPDATE Proiecte  
SET Suma=Suma*1.20  
WHERE NumePr="Proiect1";
```

### **Instrucțiunea DELETE**

Instrucțiunea permite ștergerea unor rânduri ale unui tabel. Forma ei este:

```
DELETE FROM NumeTabel  
[WHERE Conditie]
```

Pentru a șterge rândul corespunzător proiectului cu numele "Proiect1" din tabelul **Proiecte** scriem:

```
DELETE FROM Proiecte  
WHERE NumePr="Proiect1";
```



# JDBC

În 1996, Sun, scoate pe piață prima variantă de interfață de programare (Application Programming Interface - API) a bazelor de date numită Java DataBase Connectivity sau JDBC. Aceasta permitea conectarea, interogarea și actualizarea datelor pe baza limbajului SQL (Structured Query Language).

Java și JDBC aveau, încă de pe atunci, avantajul, în comparație cu alte medii de programare al bazelor de date, independența de platformă și de producătorul SGBD-ului. Aceeași bază de date scrisă în Java se putea rula pe servere Windows sau Solaris deoarece codul Java rulează identic pe ambele platforme. Datele puteau fi transferate de exemplu de pe Microsoft SQL Server pe Oracle ca urmare a folosirii a unui limbaj comun, SQL, pentru definiția și manipularea datelor. Până atunci, aplicațiile cu baze de date se scriau într-un limbaj proprietar al producătorului de SGBD cumpărătorul fiind limitat la a lucra tot timpul cu același software.

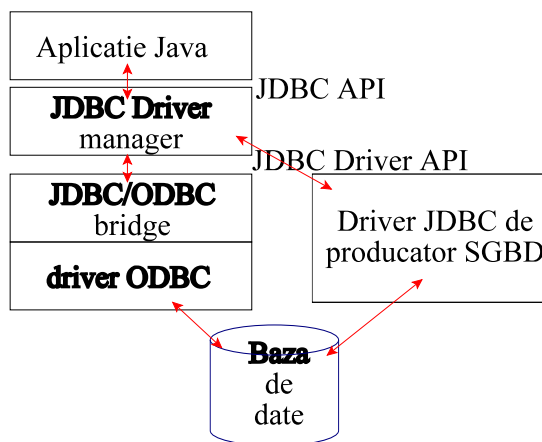
JDBC-ul suferă pe parcursul anilor mai multe extinderi însă nici după ultimul release nu are instrumente pentru dezvoltarea de interfețe grafice cu bazele de date. Aceste generatoare de formulare, interogări sau rapoarte sunt dezvoltate de diferite firme specializate numai pe acest domeniu.

# Arhitectura JDBC

Producătorii de SGBD-uri au fost de acord foarte repede ca Sun să folosească un protocol de rețea unic pentru accesul la baze de date, cu condiția ca acel protocol să fi a lor. Cum piața avea deja producători de SGBD-uri mulți și serioși Sun era în impas. Norocul vine din partea producătorilor de SGBD-uri care se pun de acord cu privire asupra faptului că Sun să creeze API-uri pure de Java pentru acces prin SQL la bazele de date împreună cu un manager de drivere care să încarce drivere externe, fabricate de firme terțe sau de către producătorii SGBD-urilor, prin care să se realizeze conexiunea la baza de date. Apoi, printr-un mecanism simplu de înregistrare al driver-elor în manager-ul de driver, toate driver-ele să respecte aceleași cerințe și opereze la fel. Ca urmare a acestor cerințe două interfețe au fost create, JDBC API pentru programatori și JDBC Driver API pentru producătorii de driver-e. Aplicațiile comunică prin API cu manager-ul de driver-e care folosește, mai departe, driver-ul dat de producător pentru a comunica cu baza de date.

Drivere-ele JDBC sunt organizate pe următoarele categorii:

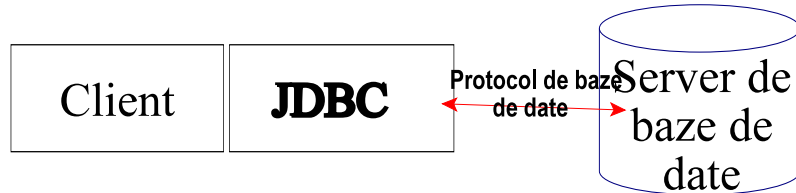
- driver tip 1: traduce JDBC în ODBC și se bazează pe driver-ul de ODBC pentru a comunica cu baza de date. Sun dă un singur driver de acest fel JDBC/ODBC bridge în JDK. Necesită configurarea corectă a driver-ului de ODBC.
- driver de tip2: este scris parțial în Java și parțial în cod nativ, comunică API-urile client ale SGBD-ului. Utilizarea unui astfel de driver obligă la instalarea unei biblioteci Java dependente de platformă.
- driver de tip 3: este o bibliotecă Java pură de tip client ce folosește un protocol independent de baza de date pentru comunicare cu baza de date utilizând o componentă server ce traduce toate cererile în cereri dependente de baza de date. Codul dependent de baza de date este stocat pe server.
- driver tip 4: o bibliotecă Java pură ce traduce cererile JDBC direct în cereri specifice protocolului bazei de date.



Majoritatea producătorilor de baze de date furnizează driver-ele de tip 3 sau 4.

# Utilizari tipice ale JDBC

Modelul tradițional este cel **client/server** cu interfață grafică complexă la client și o bază de date stocată pe un server.



Azi, arhitectura client/server este depășită fiind înlocuită cu **arhitectură pe 3 nivele** (3 tier model) sau chiar de arhitecturi mai avansate pe multi-nivele. În acest model clientul nu accesează baza de date direct.

El accesează direct doar un

nivel intermediar, de mijloc, care derulează comunicația cu baza de date. Este separată partea de prezentare, vizuală (rulată de client) de partea de logică (middle tire) și de datele fizice (din baza de date). Ca urmare a acestei separări este posibil accesul la aceleași date cu aceleși reguli de business dar de pe clienți multipli.



Comunicația între client și nivelul de mijloc se face prin HTTP dacă clientul este un navigator de web sau prin RMI când se folosește o aplicație Java. JDBC gestionează comunicația între nivelul de mijloc și baza de date. Acest model are numeroase variații, în particular Java 2 Enterprise definește o structură pentru server-ul de aplicații ce gestionează modulele de cod numite Enterprise JavaBeans ce asigură servicii ce accesare rapidă, securizată a bazei de date.

# Concepte de baza in programarea JDBC

Programarea cu JDBC nu diferă de programarea clasică cu obiecte din Java. Obiecte se creează pe bază de clase JDBC, eventual dacă este cazul se moștenesc și se extind.

## Identificarea bazei de date

Conectarea la o bază de date presupune specificarea sursei împreună cu parametrii adiționali cum sunt portul de comunicație sau valori de attribute specifice de driver (mai ales în cazul celor ODBC). URL-ul bazei de date specifică sursa de date. Printre altele aceasta trebuie să conțină numele bazei de date și locul în care se află aceasta.

```
private static final String accessDBURLPrefix =
"jdbc:odbc:Driver={Microsoft Access Driver (*.mdb)};DBQ=";

private static final String accessDBURLSuffix =
";DriverID=22;READONLY=false}";

private static final String filename = "D:\\work\\test.mdb";

filename = filename.trim();

String databaseURL = accessDBURLPrefix + filename +
accessDBURLSuffix;
```

## Realizarea conexiunii

Realizarea conexiunii se poate realiza doar dacă driver-ul JDBC pentru accesul la baza de date MS Access, care în exemplul prezentat este ODBC, este configurat corect. Apoi acesta trebuie înregistrat cu mâna prin codul:

```
Class.forName("sun.jdbc.odbc.JdbcOdbcDriver");
```

Metoda `getConnection()` întoarce obiectul `Connection`. Acesta este folosit în continuare pentru executarea de instrucțiuni SQL:

```
DriverManager.getConnection(databaseURL, "", "");
```

## Executarea unei comenzi SQL

Pentru executarea unei instrucțiuni SQL trebuie creat un obiect `Statement` prin linia de cod:

```
Statement statement = connection.createStatement();
```

Apoi, instrucțiunea SQL de executate se pune într-un String:

```
String query = "SELECT * FROM Table1";
```

Iar apoi se apelează metoda `executeQuery()`

```
ResultSet resultSet = statement.executeQuery(query);
```

Acesta întoarce un obiect din clasa `ResultSet` care este o mulțime de date ce trebuie parcurse secvențial, rând cu rând. Codul tipic pentru o astfel de ciclare este:

```
while(resultSet.next()) {  
    //afisare unui rand din multimea de randuri  
}
```

Codul funcțional pentru această afișare este:

```
while(resultSet.next()) {  
    out="";  
    for(int i=1;i<=resultSet.getMetaData().getColumnCount();++i)  
        out+=resultSet.getString(i) + " ";  
    System.out.println(out);  
}
```

Dacă comanda SQL execută acțiuni cu sunt `INSERT`, `UPDATE` sau `DELET` sau comenzi de definiție a datelor (`CREATE TABLE` sau `DROP TABLE`) atunci în locul lui `executeQuery()` se execută metoda `executeUpdate()`. Ea întoarce numărul de rânduri afectate de comandă.

## **C13- Întrebari**

1. Care este motivul utilizarii JDBC in accesarea bazelor de date?
2. Ce este JDBC?
3. Ce rol are obiectul Connection?
4. Pentru ce fel de interogari se foloseste clasa Statement?