**CURS 1**

**Baza de date**

* ansamblu strucutrat de date coerente, fara redundanta inutila, astfel incat acestea pot fi prelucrate eficient de mai multi utilizatori intr-un mod ocncurent
* colectie de date persistente, care sunt folosite de catre sistemele de aplicatii ale unei anumite “intreprinderi”
* datele din BD: integrate sau partajate

**“Intreprindere”/Enterprise**

* reguli proprii de functionare
* multime de date referitoare la modul sau de operare

**Sistem de Gestiune a Bazelor de Date (SGBD/DBMS)**

* un produs software care asigura interactiunea cu o baza de date
  + definirea
  + consultarea
  + actualizarea datelor
* controleaza si trateaza cererile de acces la baza de date
* structura unui SGBD
  + complexitate variabila
  + nivel real de functionalitate difera de la produs la produs
  + 5 clase de module:
    - Programe de GBD => realizeaza accesul fizic la date
    - Pt tratarea LDD => permit traducerea unor inf in obiecte ce pot fi exploatate procedural/neprocedural
    - Pt tratarea LMD => permit utilizatorilor inserarea, stergerea, reactualizarea sau consultarea informatiei din BD
    - De control => permit controlul programelor de aplicatie, asigurarea, confidentialitatii si integritatii datelor, rezolvarea unor probleme de concurenta, recuperarea inf in cazul unor avarii sau defectiuni
* obiectivele fundamentale:
  + independenta fizica => indp. structurilor de stocare in raport cu strcuturile de date din lumea reala
  + independenta logica => fiecare grup de lucru poate sa cunoasca doar o parte a semanticii datelor, sa vada doar o submultime a datelor si numai sub forma in care le doreste
  + prelucrarea datelor de catre neinformaticieni
  + administrarea centralizata a datelor
  + coerenta datelor => informatia trebuie sa satisfaca constranageri statice/dinamice, locale/generale
  + partajabilitatea datelor
    - aplicatiile partaja date din BD in timp si simultan
    - o aplicatie poate folosi dtae ca si cum ar fi singura care le utilizeaza, fara a sti ca alta aplicatie, concurent, le poate modifica
  + neredundanta datelor
    - administrarea coerenta a atelor trebuie sa asigure neduplicarea fizica a datelor
    - pt a realiza performante referitoare la timpul de acces la date si raspuns la solicitarile utilizatorilor, se accepta o anumita redundanta a datelor
  + securitatea si confidentialitatea datelor
    - datele trebuie protejate de accesse neautorizate/rau intentionate
    - exista mecanisme ce permit identificarea si autentificarea utilizatorilor
    - exista proceduri de acces autorizat ce depind de date si de utiliz.
    - Sistemul de gestiune trb sa asigure securitatea fizica si logica a informatiei si sa garanteze ca numai utiliz. autorizati pot efectua operatii corecte asupra BD
* Limitari/dezavantaje:
  + Complexitatea si dimensiunea sistemelor pot sa creasca considerabil
  + Costul, cheltuielile periodice de intretinere
  + Costuri aditionale pt elem hardware
  + Costul conversiei aplicatiilor existente
  + Impactul unei defectiuni

**Dictionarul datelor/Data dictionary (DD)**

* metabaza de date, “date despre date”
* strcuturat si administrat ca o baza de date
* catalog de sistem
* poate fi interogat ca orice alta baza de date
* contine:
  + descrierea tuturor obiectelor unei baze de date
  + starea acestor obiecte
  + diverse constrangeri de securitate si integritate etc.

**Administratorul baze de date/Data base administrator (DBA)**

* persoana/grup de persoane ce rapsund de ansamblul activitatilor legate de baza de date:
  + analiza
  + proiectare
  + creeaza BD reala
  + implementare a elementelor tehnice de control
  + exploatare
  + intretinere => asigura functionarea sistemului si monitorizeaza performantele acestuia
  + furnizeaza diverse servicii tehnice etc.
* atributii => de proiectare, administrative, operative, de coordonare

**Arhitectura unui sistem de BD**

* BD propriu-zisa unde se stocheaza datele
* Sistemul de gestiune al BD => gestionarea si prelucrarea complexa a datelor
* DD => informatii despre date, structura lor, statistici, documentatie
* Mijloace hardware comune/specializate
* Reglementari administrative => buna functionare a sistemului
* Personal implicat:
  + Utilizatori finali
  + DBA
  + Proiectantii de BD => proiectare fizica/logica (a design-ului)
  + Administratorul datelor (DA)
    - Manager care decide ce date trebuie stocate in BD
    - Stabileste regulile de intretinere si de tratare a datelor dupa stocare

**Proiectare logica**

* Proiectare conceptuala a BD
* Presupune cunoastere completa si amanuntita a:
  + Modelului real
  + Regulilor de functionare
* Modelul creat e independent de programele de aplicatii si de limbaje de programare
* Spre un model de date => relational, orientat obiect, ierarhic

**Proiectare fizica**

* Transpune modelul logic intr-un set de tabele supuse unor constrangeri, pt care se selecteaza structuri de stocare si metode de acces specifice, astfel incat sa se asigure performante, se iau masuri privind securitatea datelor
* Preia modelul logic de date => transpus intr-un set de tabele cu constrangeri
* Stabileste cum va fi realizat fizic
* Cunoasterea functionalitatilor SGBD => avantaje/desavantaje ale alternativelor

**Cerinte minimale BD**

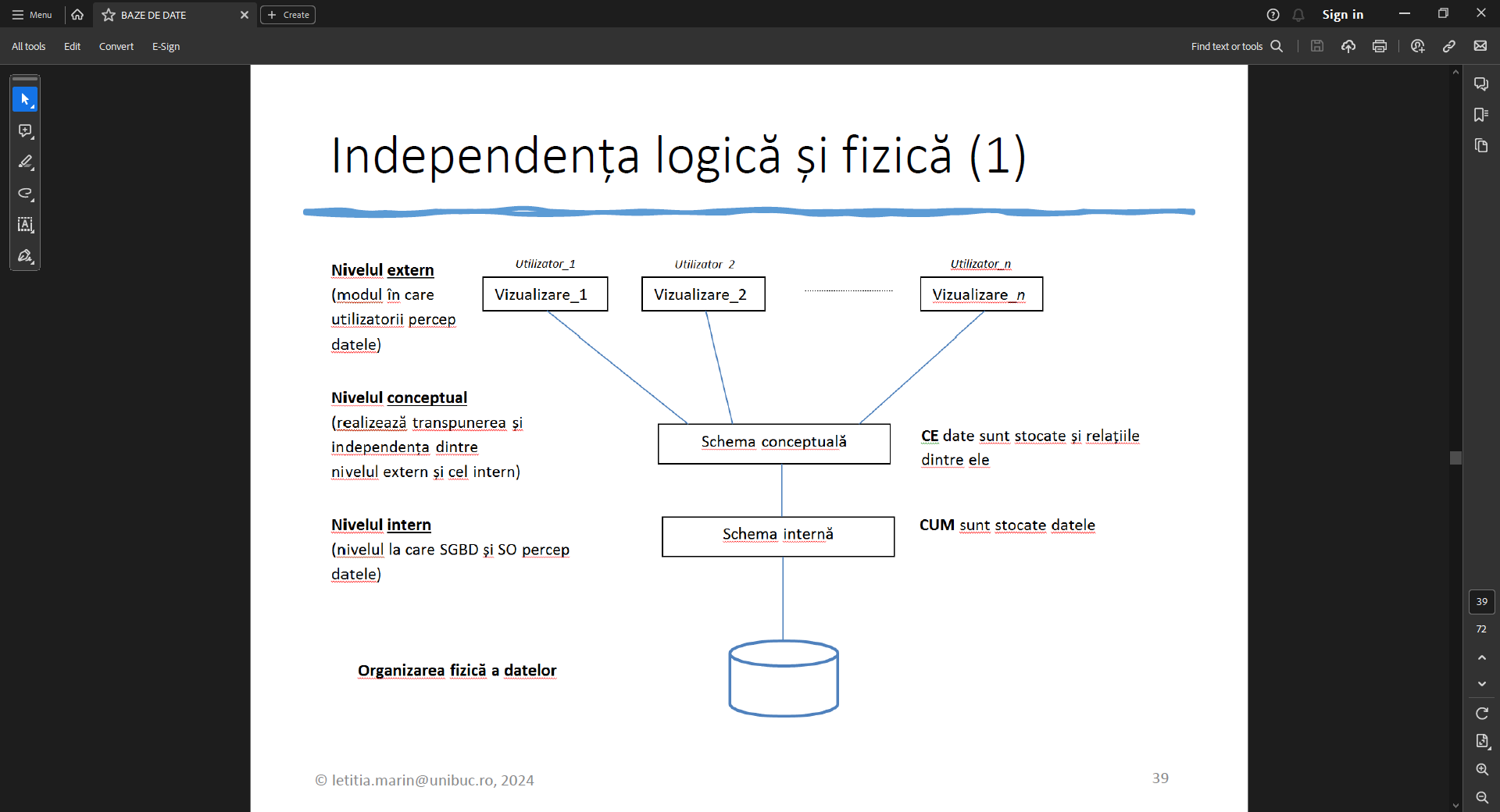
* Redundanta minima in date
* Furnizarea in timp util a informatiilor solicitate (timpul de raspuns la o interogare)
* Asigurarea unor costuri minime in prelucrare si intretinerea informatiei
* Capacitatea de a satisface, cu aceleasi date, necesitati informationale ale unui nr. mare de utilizatori
* Flexibilitate = osibilitatea de adaptare la cerinte noi, raspunsuri la interogari neprevazute initial
* Sincronizare = exploatarea simultana a datelor de catre mai multi utilizatori
* Confidentialitate = asigurarea securitatii datelor prin mecanisme de protectie importiva accesului neautorizat
* Integritate = inglobarea unor facilitati destinate validarii datelor si recuperarii lor in cazul unor deteriorari accidentale, garantarea ca datele din baza de date sunt corecte
* Compatibilitate si expandabilitatea = posibilitatea de valorificare a eforturilor anterioare si anticiparea nevoilor viitoare
* Permisivitatea, prin ierarhizarea datelor dupa criteriul freecventei acceselor, a unor reorganizari (eventual dinamice) care sporesc performantele bazei

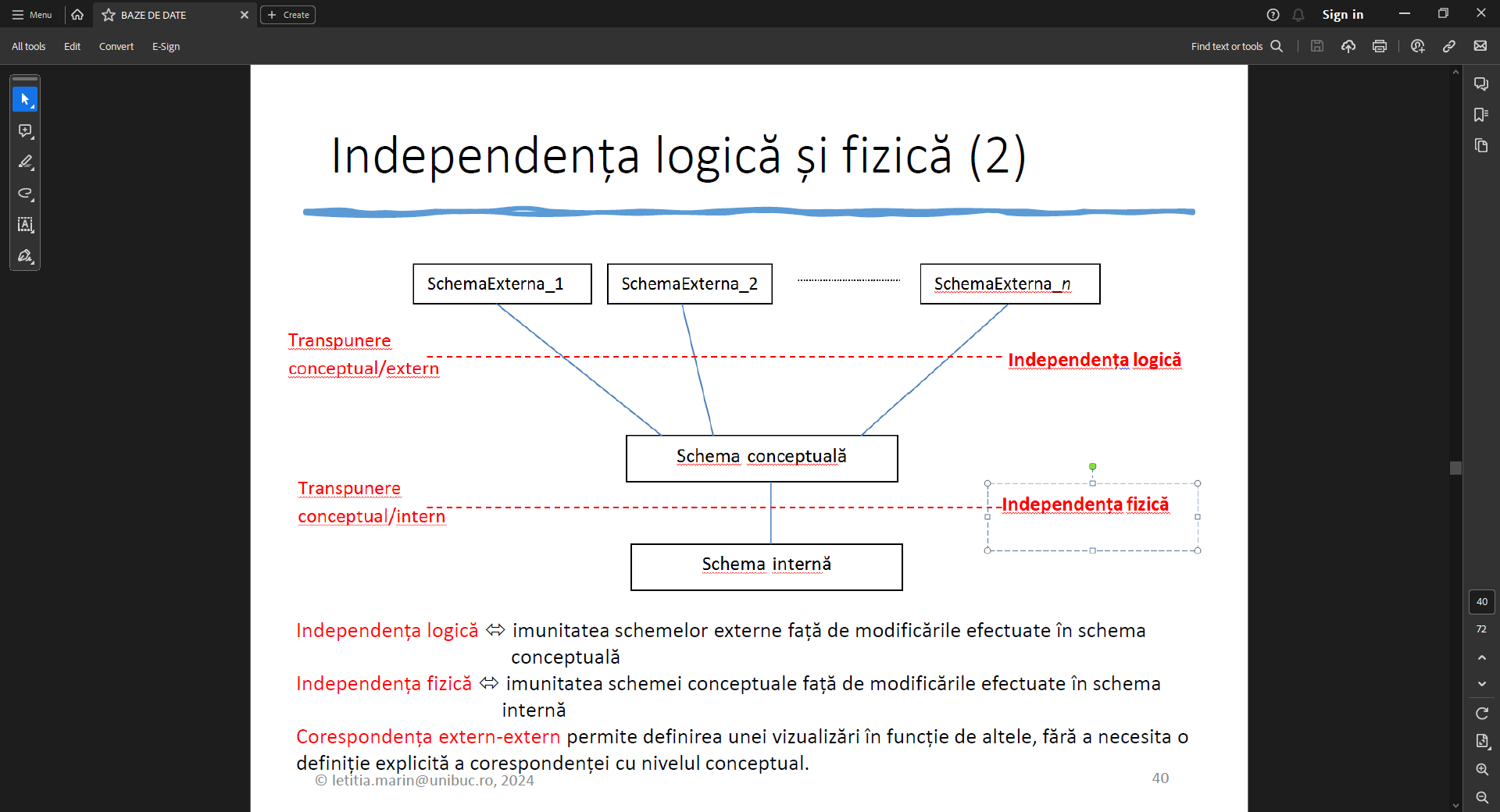
**Niveluri de abstractizare si de perceptie a datelor**

* Datele exista doar la nivel fizic, iar celelalte trei niveluri reprezinta virtualizari ale acestora.
* 4 niveluri de abstractizare:
  1. Nivel fizic/intern => schema fizica a datelor (bit, octet, adresa)
  2. Nivel conceptual => schema conceptuala a datelor (articol, inregistrare, zona) si reprezinta viziunea programatorilor de sistem asupra datelor
  3. Nivel logic => una din schemele logice posibile ale datelor si reprezinta viziunea programatorului de aplicatie asupra datelor
  4. Nivel virtual/extern => reprezinta viziunea utilizatorului final asupra datelor

**Independenta datelor**

* O modificare a structurii fizice nu va afecta aplicatia
* Reciproc, modificari ale aplicatiei vor lasa nealterata strcutura fizica de date
* 2 tipuri principale:
  + Independenta fizica => posibilitatea modificarii schemei fizice a datelor fara ca aceasta sa implice modificarea schemei conceptuale, logice si a programelor de aplicatie
  + Independenta logica
    - posibilitatea modificarii schemei conceptuale a datelor fara ca aceasta sa implice modificarea schemei logice si a programelor de aplicatie;
    - utilizatorul are iluzia ca este singurul beneficiar al unor date pe care, in realitate, le foloseste in comun cu alti utilizatori
* Independenta fata de strategiile de acces => permite programului sa precizeze data pe care doreste sa o acceseze, dar nu modul cum acceseaza acesat data
  + SGBD stabileste drumul optim de acces la date





* Independenta fizica si logica => adoptarea unei arhitecturi de BD

**Limbaje pentru BD**

* Arhitectura de BD:
  + Nivel extern
    - Schema externa
      * descriere a unei parti a BD ce corespunde viziunii unui program/utilizator
      * modelul extern e dependent de limbajul utilizat
      * permite asigurarea unui aspect al securitatii datelor
  + Nivel intern
    - Strucura interna de stocare a datelor
    - Schema interna => descrierea datelor unei BD sub forma in care sunt stocate in memoria calculatorului; descriu BD
  + Nivel central/conceptual
    - structura canonica a datelor ce caracterizeaza procesul de modelat
    - schema conceptuala => permite definirea tipurilor de date (simple/compuse), a regulilor; descriu BD
* Arhitectura functionala de referinta => propusa de ANSI/X3/SPARC, axata pe DD:
  + descrierea datelor => compozitia DD
  + prelucrarea datelor => interogarea si reactualizarea BD
* Are la baza 2 niveluri: schema si vizualizarea

**Limbaje pentru BD**

1. Limbaje pentru definirea datelor/ Data description language (LDD)

* Definirea entitatilor si a tributelor acestora
* Precizarea relatiilor dintre date si a strategiilor de acces la ele
* Stabilirea criteriilor diferentiate de confidentialitate si de validare automata a datelor utilizate

1. Limbaje pt prelucrarea datelor/ Data manipulation language (LMD)
   * Permit aduagarea, modificarea sau stergera datelor
   * O comanda are urm structura: operatia -> criterii de selctie -> mod de acces secvential/indexat/etc.
   * 2 tipuri:
     + Procedurale => cum se obtine rezultatul unei comenzi LMD
     + Neprocedurale => descriu doar datele ce vor fi obtinute si nu modalitatea
2. Limbaje pt controlul datelor/Data control language
   * Asigura confidentialitatea si integritatea datelor
   * Salvarea informatiei in cazul unor defectiuni
   * Rezolvarea unor probleme de ocncurenta

**Evolutia BD**

* 3 generatii:
  + Sisteme ierarhice
  + Sisteme relationale
  + Sisteme avasnate => orientatea obiect, relationale or. ob. Deductive, decizionale, distribuite, magazii de date
* Datele sunt reprezentate prin legaturi ierarhice (arbore)/tip graf
* Limbajul de prelucrare a datelor impune programatorului sa specifice drumurile de acces la date
* Modelul relational => trateaza entitatile ca pe niste relatii algebrice
* BD actuale sunt acoperite majoritatea de sisteme relationale
* BD relationale:
  + Structuri de date simple, intuitive (NU/ nimic dinamice)
  + Inexistenta pointerilor vizibili pt utilizatori
  + Constrangeri de integritate
  + Operatori pt relatii ce permit deifnirea, cautarea si reactualizarea datelor
  + Nu realizeaza gestiunea datelor distribuite si cea a cunostintelor
  + Nu modeleaza grafuri, liste etc.

**SGBDOO**

* Tehnica limbajelor orientate obiect + BD => BD orientate obiect
* Avantaje:
  + Modelare superioara a informatiei
  + Posibilitati superioare de deductie (ierarhie de clase, mostenire)
  + Luarea in considerare a aspectelor dinamice
  + Integrarea descrierii structurale si comportamentale
* Dezavantaje:
  + Gestiunea obiectelor complexe este mai dificila
  + Sume uriase investite de utilizatori

**SGBDROO**

* Simplitatea modelului relational + puterea tehnologiei orientate obiect => BD relationale orientate obiect
* Adaugarea caracteristiclor obiectuale necesare sau pronind de la un sistem orientat obiect si adaugand caracteristicile relationale

**BD deductiv**

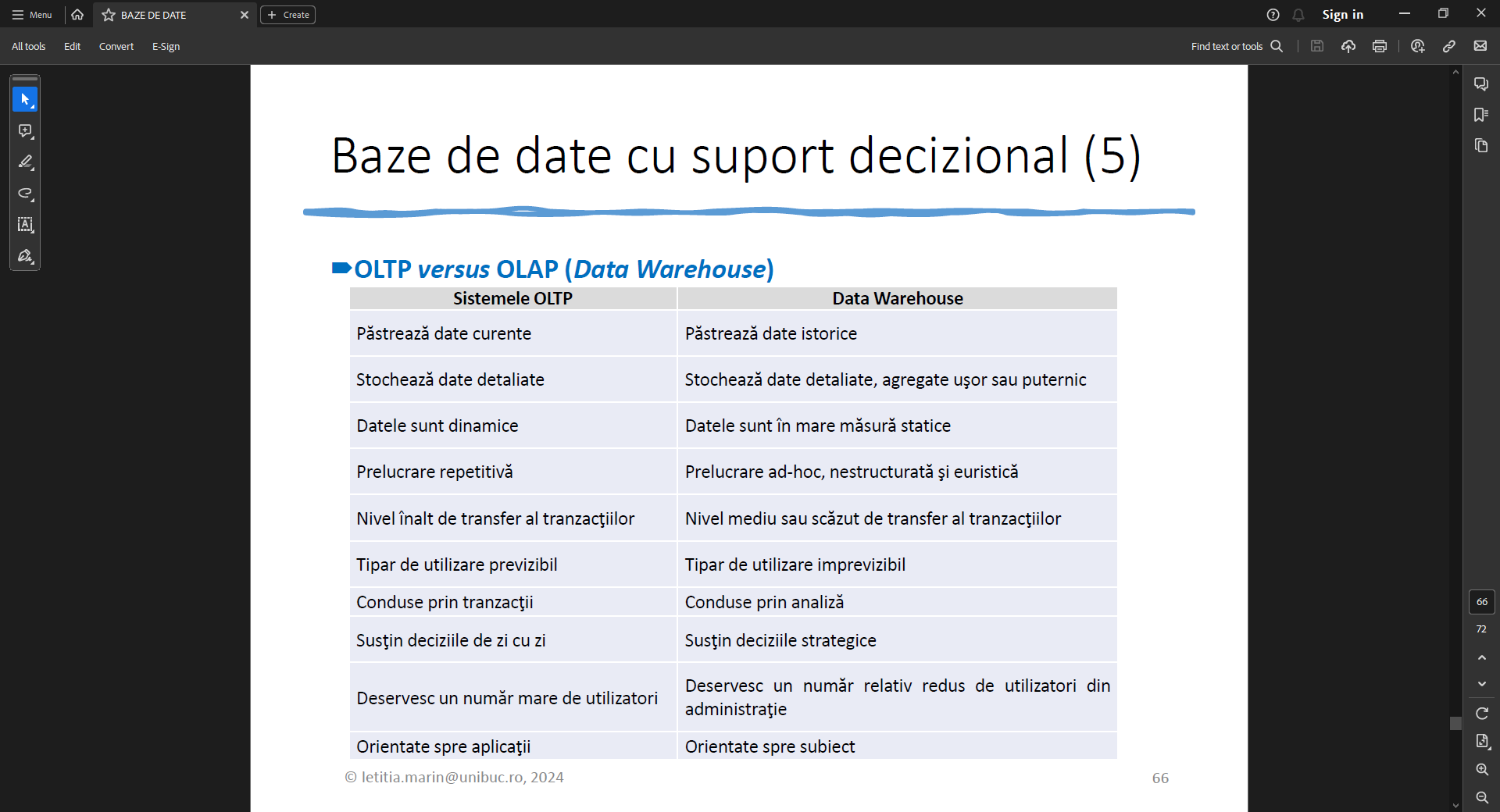
* O relatie este o multime de inregistrari ce reprezinta fapte
* Cunostintele sunt asertiuni generale si abstracte asupra faptelor, permit deducerea faptelor noi
* SGBD relational suporta o forma limitat de cunostine=constrangerile de integritate, restul trb integrate
* Prin programare logica se gestioneaza cunostinte relationale relativ la BD
* SGBD deductiv are:
  + LDD ce permite definirea structurii predicatelor sub forma de relatii si constrangeri de int asociate
  + Limbaj de prelucrare a datelor
  + Limbaj de reguli de deductie ce permite contruirea predicatelor derivate

**BD distribuite (BDD)**

* Sistem distribuit = anasmablu de masini ce sunt interconectate printr-o retea de comunicatie si utilizate intr-un scop global
* Administrarea si prelucrarea datelor distribuite situate pe dif calculatoare si exploatate de sisteme eterogene
* Multimea de BD e exploatata de utilizator ca si cum e doar una
* Programul de aplicatie ce exploateaza BDD poate avea acces la date rezidente pe mai multe masini, fara ca programatorul sa cunoasca localizarea datelor
* Modelul relational => instrumentul principal prin care se realizeaza prelucrarea datelor distribuite

**BD cu suport decizional**

* Cantitate mare de inf ale companiilor pastrata in tabele istorice
* Analiza date istorice => informatii despre tendinte si evolutii care ar putea interesa compania
* Intampina nevoile sistemelor operationale a caror natura e inerent tranzactionala
* Sistemele tranzactionale controleaza la un mom dat o singura tranzactie
* Sistemul operational tipic opereaza cu evenimente predefinite si necesita acces rapid la date
* BD optimizata => Data Warehouse (magazie de date)
* Procesare analitica/On line analytical processing (OLAP) => principiul urmat
* Procesare tranzactionala/On line transactional proccessing(OLTP) => principiul pe care se bazeaza sistemele tranzactionale
* Intergoraile de BD analitice sunt ad-hoc, nu unt predefinite
* Data mining/knowledge discovery in database (KDD) => procesul de descoperire a inf utile din BD pt a construi DSS (Decision support system)



**Arhitectura multitier/mai multe niveluri**

* Contine:
  + Unul/mai multi clienti care initiaza operatii
  + Unul/mai multe servere de aplicatii care exxecuta parti ale operatiilor
  + Un server de BD care stocheaza datele folosite de operatii
* Privilegiile serverului de aplicatii sunt limitate pt a preveni executia operatiilor nedorite sau inutile in timpul unei operatii client
* Clientul (un browser web/proces user) => cererea pt BD
* Serverul de aplicatii => interfata dintre clienti si serverul BD, asigurand accesul la informatii, include un nivel aditional de securitate
* Serverul BD => pune la dispozitia serverului de palicatii inf pt solutionarea operatiilor facute de client