

O projektovanju baza podataka

prema

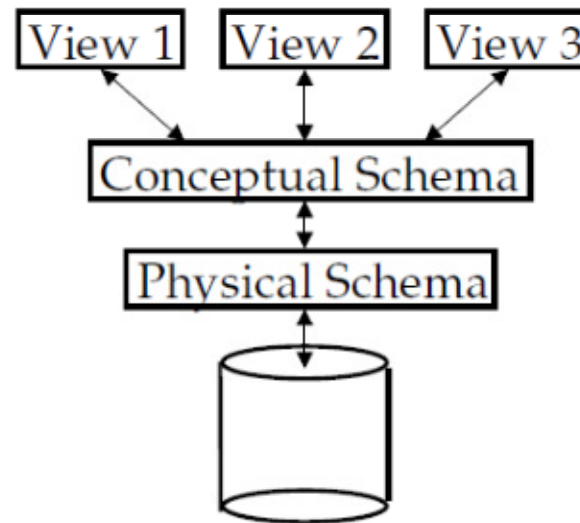
Database Management Systems

Raghu Ramakrishnan,
Johannes Gehrke

Nivoi apstrakcije

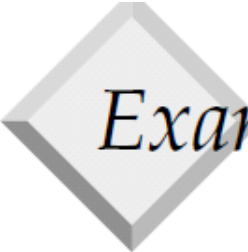
Levels of Abstraction

- ❖ Many views, single conceptual (logical) schema and physical schema.
 - Views describe how users see the data.
 - Conceptual schema defines logical structure
 - Physical schema describes the files and indexes used.



☛ Schemas are defined using DDL; data is modified/queried using DML.

Nivoi apstrakcije



Example: University Database

- ❖ Conceptual schema:
 - *Students(sid: string, name: string, login: string, age: integer, gpa: real)*
 - *Courses(cid: string, cname: string, credits: integer)*
 - *Enrolled(sid: string, cid: string, grade: string)*
- ❖ Physical schema:
 - Relations stored as unordered files.
 - Index on first column of Students.
- ❖ External Schema (View):
 - *Course_info(cid: string, enrollment: integer)*

Projektovanje (dizajn) baze podataka

- Kako korisnik opisuje instituciju
 - (firmu, preduzeće) realnog sveta, npr. univerzitet,
 - u terminima podataka uskladištenih u bazi podataka?
- Koje faktore treba uzeti u razmatranje?

Projektovanje (dizajn) baze podataka: alati

- **Alati za projektovanje baze podataka**
 - **Proizvođači RSUBP, ostali**
 - Npr, IBM Rational Rose
<http://www-03.ibm.com/software/products/us/en/ratirosefami/>
 - Oracle: <http://www.oracle.com/tools>
 - DBDesigner 4- MySQL dizajn alat
<http://www.fabforce.net/dbdesigner4/>
 - Sybase: [http://www.sybase.com/products/application tools](http://www.sybase.com/products/application%20tools)

Projektovanje (dizajn) baze podataka: koraci

- **6 koraka**
- ER model je najvažniji za prva 3 koraka
- **(1) Analiza zahteva: Šta korisnik želi od baze podataka**
 - koje podatke treba čuvati u bazi podataka,
 - koje aplikacije treba izgraditi nad njima i
 - koje su najčešće operacije i koji su zahtevi za njihove performanse
 - Obično neformalni proces – diskusija s korisničkim grupama, analiza radnog okruženja, očekivanih promena, dokumentacije o postojećim aplikacijama koje bi trebalo podržati bazom podataka
 - Postoje metodologije i alati za podršku ovog procesa - npr. JAD (***Joint application design***), intervju, upitnik... / RAT, RATSU (**RATSU -- Requirements Analysis Tool with Synthesis**);
UML

Projektovanje baze podataka: ER-koraci

- **(2) Konceptualno projektovanje baze podataka**
 - Prikupljena informacija se koristi za opis, visokog nivoa apstrakcije, podataka koji će se čuvati i ograničenja za te podatke.
 - Obično se koristi ER model podataka
- **(3) Logičko projektovanje baze podataka**
 - Izbor SUBP za implementaciju projekta baze podataka
 - Konverzija konceptualne sheme u shemu baze podataka odgovarajućeg modela (izabranog SUBP)
 - Kada je izabrani SUBP – RSUBP – konverzija ER sheme u shemu relacione baze podataka
 - Rezultat: konceptualna shema (zove se i logička shema) u RM

Projektovanje baze podataka: posle-ER

- ER dijagram je samo aproksimacija opisa podataka
- Subjektivna evaluacija informacije prikupljene u koraku 1
- Dalja analiza – ***profinjenje logičke sheme***
- Kriterijumi performansi – ***projektovanje fizičke sheme***
- Pitanja bezbednosti: ***pristup podacima i zabrane***

Projektovanje baze podataka: normalizacija

- **(4) Profinjenje sheme**
 - Analiza skupa relacija u relacionoj shemi – otkrivanje potencijalnih problema
 - Za razliku od prethodnih koraka, za profinjenje sheme postoji moćna teorija
 - Teorija normalizacije

Projektovanje baze podataka: fizičko projektovanje

- **(5) Fizičko projektovanje baze podataka**
 - Analiza tipičnog radnog opterećenja (workload)
 - Dalje profinjenje projekta bp koje obezbeđuje kriterijume performansi
 - Izgradnja nekih indeksa, nekih sortova (klastera) **ALI I**
 - Reprojektovanje nekih delova sheme bp dobijene u prethodnim koracima

Projektovanje baze podataka: bezbednost

- **(6) Projektovanje bezbednosti:**
 - Identifikuju se različite korisničke grupe i uloge koje korisnici imaju
 - Na primer, menadžer proizvodnje, agent podrške potrošača i sl.
 - Identifikovati delove baze podataka kojima svaki korisnik / uloga može odnosno ne sme da pristupi
 - Preduzeti mere za obezbeđivanje predviđenog pristupa
 - SUBP ima nekoliko mehanizama podrške ovog koraka

Projektovanje baze podataka: dodatne radnje

- Navedenih 6 koraka predstavlja samo *klasifikaciju* poslova
- Potrebno je još i:
 - (obično) neko “**štelovanje**” (*tuning*): svih 6 koraka se ponavlja, preklapa, do zadovoljavajućeg dizajna
 - **Implementacija** dizajna
 - Dizajn i implementacija **aplikacionih nivoa** nad SUBP.
- Mogu da utiču na korekcije osnovnog projekta bp

Projektovanje baza podataka: studija slučaja – Internet prodavnica

- DBDudes Inc., a well-known database consulting firm, has been called in to help Barnes and Noble (B&N) with their database design and implementation.
- B&N is a large bookstore specializing in books on horse racing, and they've decided to go online.
- DBDudes first verify that B&N is willing and able to pay their steep fees and then schedule a lunch meeting - billed to B&N, naturally - to do requirements analysis.

Studija slučaja – Internet prodavnica

Analiza zahteva

- Vlasnik B&N iznosi zahtev:

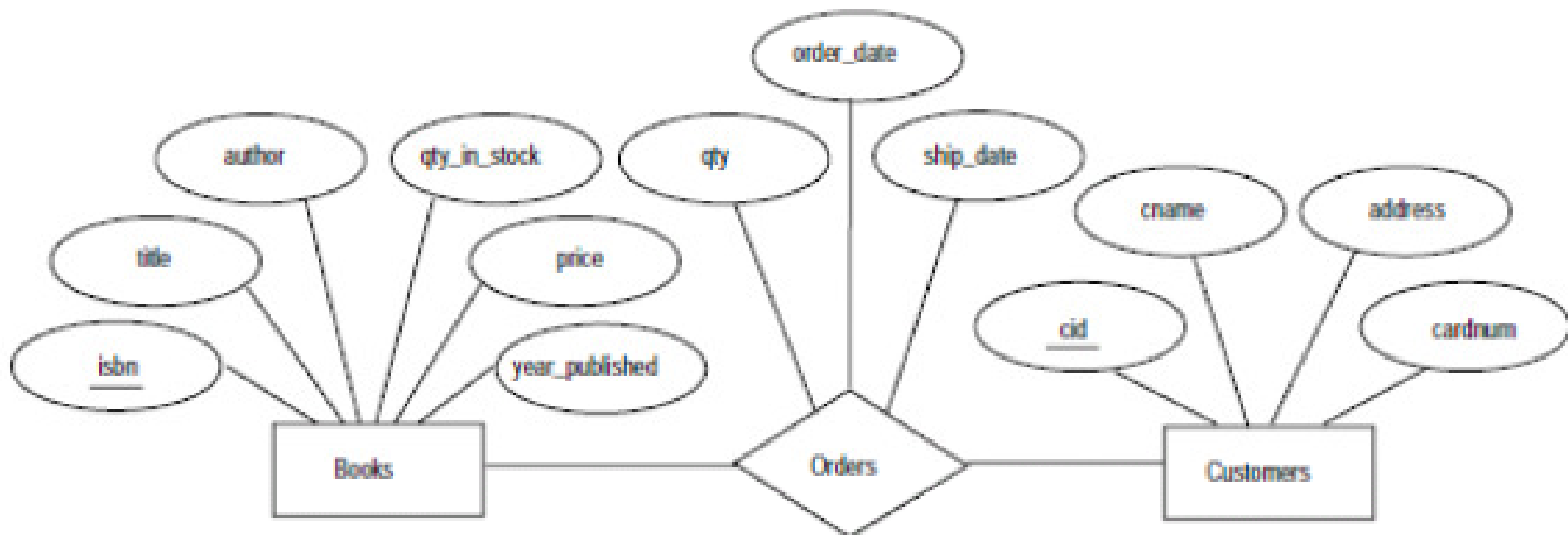
“I would like my customers to be able to browse my catalog of books and to place orders over the Internet. Currently, I take orders over the phone. I have mostly corporate customers who call me and give me the ISBN number of a book and a quantity. I then prepare a shipment that contains the books they have ordered. If I don't have enough copies in stock, I order additional copies and delay the shipment until the new copies arrive; I want to ship a customer's entire order together. My catalog includes all the books that I sell. For each book, the catalog contains its ISBN number, title, author, purchase price, sales price, and the year the book was published. Most of my customers are regulars, and I have records with their name, address, and credit card number. New customers have to call me first and establish an account before they can use my Web site.

On my new Web site, customers should first identify themselves by their unique customer identification number. Then they should be able to browse my catalog and to place orders online.”

Studija slučaja – Internet prodavnica

Konceptualno projektovanje

- Inicijalni opis DBDudes visokog nivoa – ER dijagram:



- interpretacija: knjige (Books) i korisnici (Customers) modeliraju se kao entiteti i povezuju se narudžbama (Orders); to je odnos; ima attribute; attribute shp_date ima vrednost NULL pre slanja narudžbe

Studija slučaja – Internet prodavnica

Konceptualno projektovanje

- Prva interna recenzija (recenzent “Dude 2”):
 - *Dude 2:* What if a customer places two orders for the same book on the same day?
 - *Dude 1:* The first order is handled by creating a new Orders relationship and the second order is handled by updating the value of the quantity attribute in this relationship.
 - *Dude 2:* What if a customer places two orders for different books on the same day?
 - *Dude 1:* No problem. Each instance of the Orders relationship set relates the customer to a different book.
 - *Dude 2:* Ah, but what if a customer places two orders for the same book on different days?
 - *Dude 1:* We can use the attribute order date of the orders relationship to distinguish the two orders.
 - *Dude 2:* Oh no you can't. The attributes of Customers and Books must jointly contain a key for Orders. So this design does not allow a customer to place orders for the same book on different days.
 - *Dude 1:* Yikes, you're right. Oh well, B&N probably won't care; we'll see.

Studija slučaja – Internet prodavnica

Logičko projektovanje

- DBDudes preslikavaju ER dijagram u RM:

```
CREATE TABLE Books ( isbn CHAR(10),  
                    title CHAR(80),  
                    author CHAR(80),  
                    qty_in_stock INTEGER,  
                    price REAL,  
                    year published INTEGER,  
                    PRIMARY KEY (isbn))
```

```
CREATE TABLE Orders ( isbn CHAR(10),  
                      cid INTEGER,  
                      qty INTEGER,  
                      order date DATE,  
                      ship date DATE,  
                      PRIMARY KEY (isbn,cid),  
                      FOREIGN KEY (isbn) REFERENCES Books,  
                      FOREIGN KEY (cid) REFERENCES Customers )
```

```
CREATE TABLE Customers ( cid INTEGER,  
                          cname CHAR(80),  
                          address CHAR(200),  
                          cardnum CHAR(16),  
                          PRIMARY KEY (cid)  
                          UNIQUE (cardnum))
```

Studija slučaja – Internet prodavnica

Logičko projektovanje

U pokušaju da prevaziđe neplanirano ograničenje, *Dude 1* predlaže proširenje ključa odnosa Order koje nije u skladu sa ER dijagramom:

```
CREATE TABLE Orders ( isbn CHAR(10),  
    ...  
    PRIMARY KEY (isbn,cid, ship_date),  
    ... )
```

Studija slučaja – Internet prodavnica

Logičko projektovanje

- Prezentacija B&N:
- Novi zahtevi:
 - “Customers should be able to purchase several different books in a single order. In addition, they could place more than one order per day, and they want to be able to identify the orders they placed.”
- Diskusija o načinu slanja knjiga:
- B&N: “As soon as we have enough copies of an ordered book we ship it, even if an order contains several books.”
- DBDudes: novi atribut *ordernum* u relaciju Orders
- CREATE TABLE Orders (*ordernum* INTEGER,
 isbn CHAR(10),
 cid INTEGER,
 qty INTEGER,
 order_date DATE,
 ship_date DATE,
 PRIMARY KEY (*ordernum*, *isbn*),
 FOREIGN KEY (*isbn*) REFERENCES Books
 FOREIGN KEY (*cid*) REFERENCES Customers)

Studija slučaja – Internet prodavnica

Profinjenje sheme

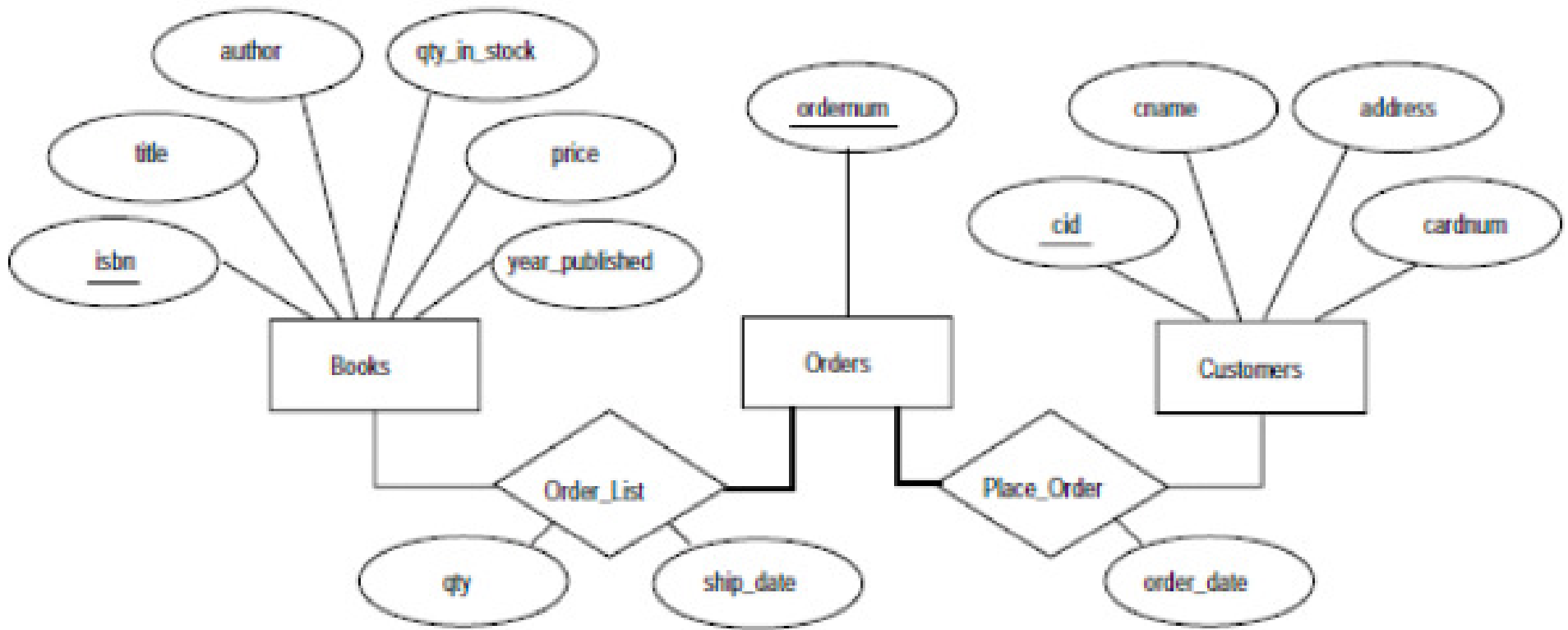
- DBDudes analiziraju moguće redundanse
- Relacija Books u BCNF: samo jedan ključ (*isbn*) i nema drugih FZ
- Relacija Customers ima ključ *cid*, i FZ *cardnum* → *cid*, pa je i *cardnum* ključ. Nema drugih FZ pa je i relacija Customers u BCNF.
- Relacija Orders ima ključ (*ordernum*, *isbn*) i važe FZ *ordernum* → *cid*, i *ordernum* → *order_date*, pa ona nije ni u 3NF (zašto? Da li je u 2NF?)
- Relaciju Orders dekomponuju u dve relacije, obe u BCNF:
Orders(*ordernum*, *cid*, *order_date*) i
Orderlists(*ordernum*, *isbn*, *qty*, *ship_date*)
- Dekompozicija je *lossless join* i čuva polazne FZ

Studija slučaja – Internet prodavnica

Profinjenje sheme

- CREATE TABLE Orders (ordernum INTEGER,
cid INTEGER,
order_date DATE,
PRIMARY KEY (ordernum),
FOREIGN KEY (cid) REFERENCES Customers)
- CREATE TABLE Orderlists (ordernum INTEGER,
isbn CHAR(10),
qty INTEGER,
ship_date DATE,
PRIMARY KEY (ordernum, isbn),
FOREIGN KEY (isbn) REFERENCES Books)
- Figure A.2 shows an updated ER diagram that reflects the new design.
- Note that DBDudes could have arrived immediately at this diagram if they had made Orders an entity set instead of a relationship set right at the beginning. But at that time they did not understand the requirements completely, and it seemed natural to model Orders as a relationship set.
- This iterative refinement process is typical of real-life database design processes. As DBDudes has learned over time, it is rare to achieve an initial design that is not changed as a project progresses.

Studija slučaja – Internet prodavnica novi ER dijagram



Orders je postao entitet!

Studija slučaja – Internet prodavnica

Fizičko projektovanje

- DBDudes razmatraju očekivano **radno opterećenje**
- Vlasnik očekuje najintenzivnije pretraživanje po isbn pre naručivanja
- Naručivanje uključuje unošenje jednog sloga u tabelu Orders i jedan ili više slogova u tabelu Orderlists
- Ako ima dovoljno primeraka, priprema se pošiljka i postavlja vrednost za *ship_date*. Smanjuje se raspoloživa količina
- Nove knjige stižu od dobavljača i povećavaju raspoložive količine

- Odluke:
 - **Books:**
 - Pretraživanje po *isbn* je pretraživanje na jednakost, heširanje
 - Ažuriranje količine (*qty_in_stock*) – prvo pretraga po *isbn*; efikasna
 - Ažuriranje *qty_in_stock* često – vertikalno particionisanje:
BooksQty(*isbn*, *qty*), and
BookRest(*isbn*, *title*, *author*, *price*, *year_published*)
 - Usporava pretragu svih informacija o knjizi (spajanje) – pa se odustaje

Studija slučaja – Internet prodavnica

Fizičko projektovanje

- DBDudes očekuju česte pretrage po naslovu i autoru, pa dodaju indekse po atributima *title* i *author* – jeftini jer se naslovi i autori retko menjaju
- **Customers**
- Najčešći upiti su pretraga na jednakost sa *cid* – heš indeks po *cid*
- **Orders**
- Dva upita: unošenje novih i pretraga postojećih
- *ordernum* atribut uključen: grupišući B+ indeks (dodeljuju se redom i odgovaraju datumu, pa je sortiranje po *ordernum* istovremeno i sortiranje po *order_date*)
- Ovo je odluka DBDudes i olakšava dnevno praćenje, ali otežava pretragu svih narudžbi datog kupca, ako kupac ima puno narudžbi (ne može da postoji i grupišući indeks i po *cid*)

Studija slučaja – Internet prodavnica

Fizičko projektovanje

– Orderlist

- Uglavnom uključuje unošenja slogova i povremeno izmenu datuma pošiljke, ili
- Listanje komponenti narudžbe
- Grupišući B+ indeks po *ordernum*
- Efikasna pretraga svih stavki jedne narudžbe
- Za ažuriranje datuma, pretraga po paru (*ordernum*, *isbn*) – pomaže indeks po *ordernum*
- Indeks po paru (*ordernum*, *isbn*) bi usporio unošenje (insert) slogova
- Ostaje samo indeks po *ordernum*

Studija slučaja – Internet prodavnica

digresija: “šteloavanje” (tuning)

- Posle nekoliko meseci od uvođenja baze podataka, stiže žalba da se upiti nad narudžbama vrlo sporo obrađuju
- B&N su postali uspešni, tabele Orders i Orderlists su postale ogromne
- DBDudes uočavaju dva tipa narudžbi:
 - Kompletirane – sve knjige poslate (većina)
 - Delimično kompletirane (mali deo)
 - Većina kupaca nadgleda delimično kompletirane narudžbe
 - Horizontalno particionisanje tabela Orders i Orderlists po *ordernum*.
 - Četiri nove relacije: NewOrders, OldOrders, NewOrderlists, i OldOrderlists
 - Usporeni su samo (retki) upiti koji zahtevaju i stare i nove narudžbe, npr. sve narudžbe jednog kupca

Studija slučaja – Internet prodavnica bezbednost

- Tri grupe korisnika: kupci, zaposleni i vlasnik (naravno, i db administrator sa svim pravima)
- Vlasnik ima sve privilegije nad svim tabelama
- Kupci mogu da pretražuju tabelu Books i da izdaju narudžbe ali nemaju pristup slogovima drugih kupaca i njihovih narudžbi
- DBDudes: Ograničenje pristupa na dva načina
 - Veb stranica sa nekoliko formi – mali broj različitih zahteva (bez mogućnosti SQL interfejsa)
 - Pretraga Books po isbn, author, title
 - Dva dugmeta: lista delimično kompletiranih i lista kompletiranih narudžbi
 - Poslovna logika: održavanje *cid* (pri logovanju na veb sajt)
 - Svojstva bezbednosti SUBP-a

Studija slučaja – Internet prodavnica bezbednost kroz SUBP

- Ograničavanje pristupa prem korisničkoj grupi
- **Nalog customer** sa privilegijama:
SELECT ON Books, NewOrders, OldOrders, NewOrderlists, OldOrderlists
INSERT ON NewOrders,, NewOrderlists,
- Zaposleni mogu da dodaju knjigu katalogu, ažuriraju količinu knjiga u magacinu, izmene narudžbu kupca, ažuriraju podatke o kupcu OSIM informacije o kreditnoj kartici (ne treba ni da je vide)
- Pogled CustomerInfo
CREATE VIEW CustomerInfo (cid,cname,address)
AS SELECT C.cid, C.cname, C.address
FROM Customers C
- **Nalog employee** ima privilegije:
SELECT ON CustomerInfo, Books,
NewOrders, OldOrders, NewOrderlists, OldOrderlists
INSERT ON CustomerInfo, Books,
NewOrders, OldOrders, NewOrderlists, OldOrderlists
UPDATE ON CustomerInfo, Books,
NewOrders, OldOrders, NewOrderlists, OldOrderlists
DELETE ON Books, NewOrders, OldOrders, NewOrderlists, OldOrderlists
- Logovanje na veb sajt (cid) zahteva bezbedni protokol, npr. SSH
- Neke kompanije nude elektronsko plaćanje (druga tema)

Studija slučaja – Internet prodavnica

Aplikativni nivoi

- Potreba za upravljanjem sesijom (čuvanje *cid-a*)
- Statičke ili dinamičke veb stranice za knjige; statička zahteva izradu posebne stranice za svaku knjigu; dinamička se generiše iz standardnog obrasca (templejta)
- Povezivanje aplikacije sa SUBP
 - CGI skript
 - Kodiranje logike upravljanja sesijom; nije lako
 - Infrastruktura aplikacionog servera
 - Korišćenje funkcionalnosti aplikacionog servera
 - Preporučuju da B&N implementira obradu na strani servera korišćenjem aplikacionog servera
 - B&N odbija da plati za aplikacioni server i želi CGI skriptove
 - DBDudes razvija:
 - Vršne HTML stranice za navigaciju i razne forme za pretragu kataloga i prezentaciju rezultata
 - Logika za praćenje sesije kupca – relevantna informacija ili na serveru ili u razgledaću (npr. kolačići)
 - Skriptovi za obradu zahteva
 - U slučaju izbora aplikacionog servera, drugi skup zadataka

Projektovanje baza podataka

- Iterativni proces