

Projektiranje informacijskih sustava

SDLC faza analize - Modeliranje
podataka

Ak. god. 2011/2012

Model podataka

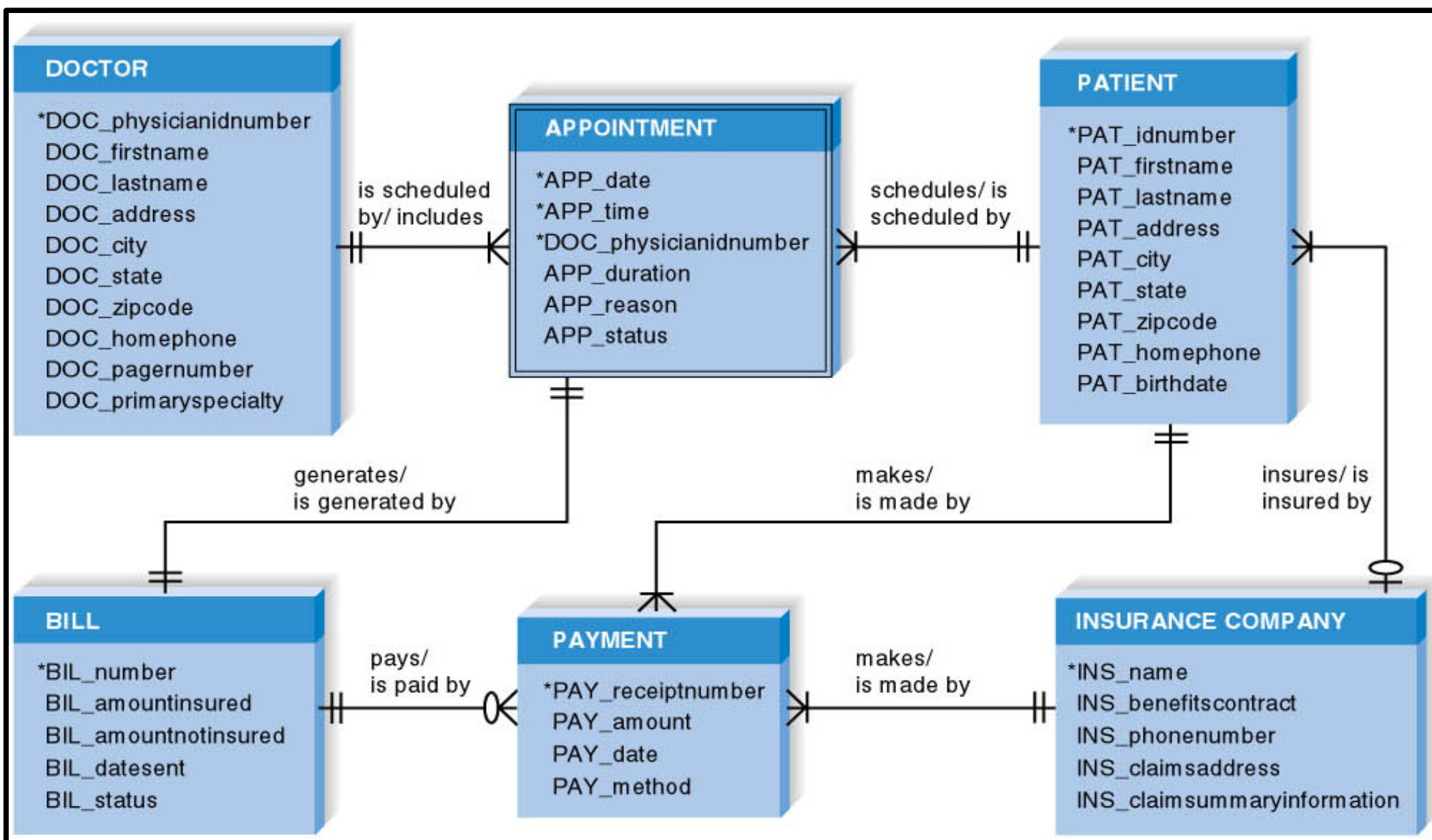
- Model podataka je formalni način predstavljanja podataka koji se koriste ili stvaraju u poslovnom procesu.
- U fazi analize razvija se logički model podataka koji se prebacuje u fizički model podataka tijekom faze dizajna.
- Logički model podataka opisuje podatke bez da sugerira kako se oni pohranjuju, kreiraju ili mijenjaju.

ERD




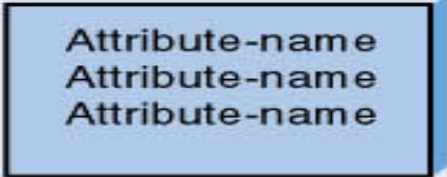

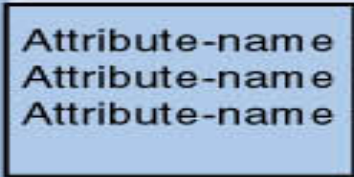

- Najčešće korištena tehnika za modeliranje podataka je ERD (*entity-relationship diagram*) (Peter Chen, 1976).
- ERD dijagram grafički prikazuje informacije koje se kreiraju, pohranjuju ili koriste u sustavu.
- Sastoji se od entiteta (koncepta) koji predstavljaju kolekciju sličnih informacija (npr. entitet osobni podaci sadrži ime, prezime, godinu rođenja,..., entitet adresa sadrži ime ulice, kućni broj, grad,...) prikazanih pravokutnikom.

ERD

- Podaci koji nemaju višestruke instance u sustavu se ne definiraju kao entiteti.
- Linije između entiteta definiraju relacije među podacima tj. postojanje nekakvog oblika odnosa među tim podacima.
- Pored relacija koje definiraju poslovna pravila nad podacima (npr. svaka osoba ima adresu), navode se i dodatne oznake koje definiraju poslovna pravila (npr. jedna osoba se može povezati sa jednom ili više adresa).



ELEMENTI ERD-a

	IDEF1X	Chen	Crow's Foot
<p>An ENTITY:</p> <ul style="list-style-type: none"> ✓ Is a person, place, or thing ✓ Has a singular name spelled in all capital letters ✓ Has an identifier ✓ Should contain more than one instance of data 	<p>ENTITY-NAME</p> 	<p>ENTITY-NAME</p> 	<p>ENTITY-NAME</p> 
<p>An ATTRIBUTE:</p> <ul style="list-style-type: none"> ✓ Is a property of an entity ✓ Should be used by at least one business process ✓ Is broken down to its most useful level of detail 	<p>ENTITY-NAME</p> 		<p>ENTITY-NAME</p> 
<p>A RELATIONSHIP:</p> <ul style="list-style-type: none"> ✓ Shows the association between two entities ✓ Has a parent entity and a child entity ✓ Is described with a verb phrase ✓ Has cardinality (1 : 1, 1 : N, or M : N) ✓ Has modality (null, not null) ✓ Is dependent or independent 	<p><u>Relationship-name</u></p>		<p><u>Relationship-name</u></p>

ERD

- Tri osnovna elementa ERD-a su:
 1. entitet,
 2. atribut
 3. relacija
- Entitet je osnovni građevni blok ERD-a. Entitet može predstavljati fizički ili logički pojam (osoba, događaj, mjesto ili stvar), tj. bilo što o čemu se skupljaju podaci (student, ispit, ...).
- Svaki entitet ima Ime i Identifikator. Ime entiteta je imenica i piše se velikim slovima. Identifikator jednoznačno određuje jednu instancu nekog entiteta.

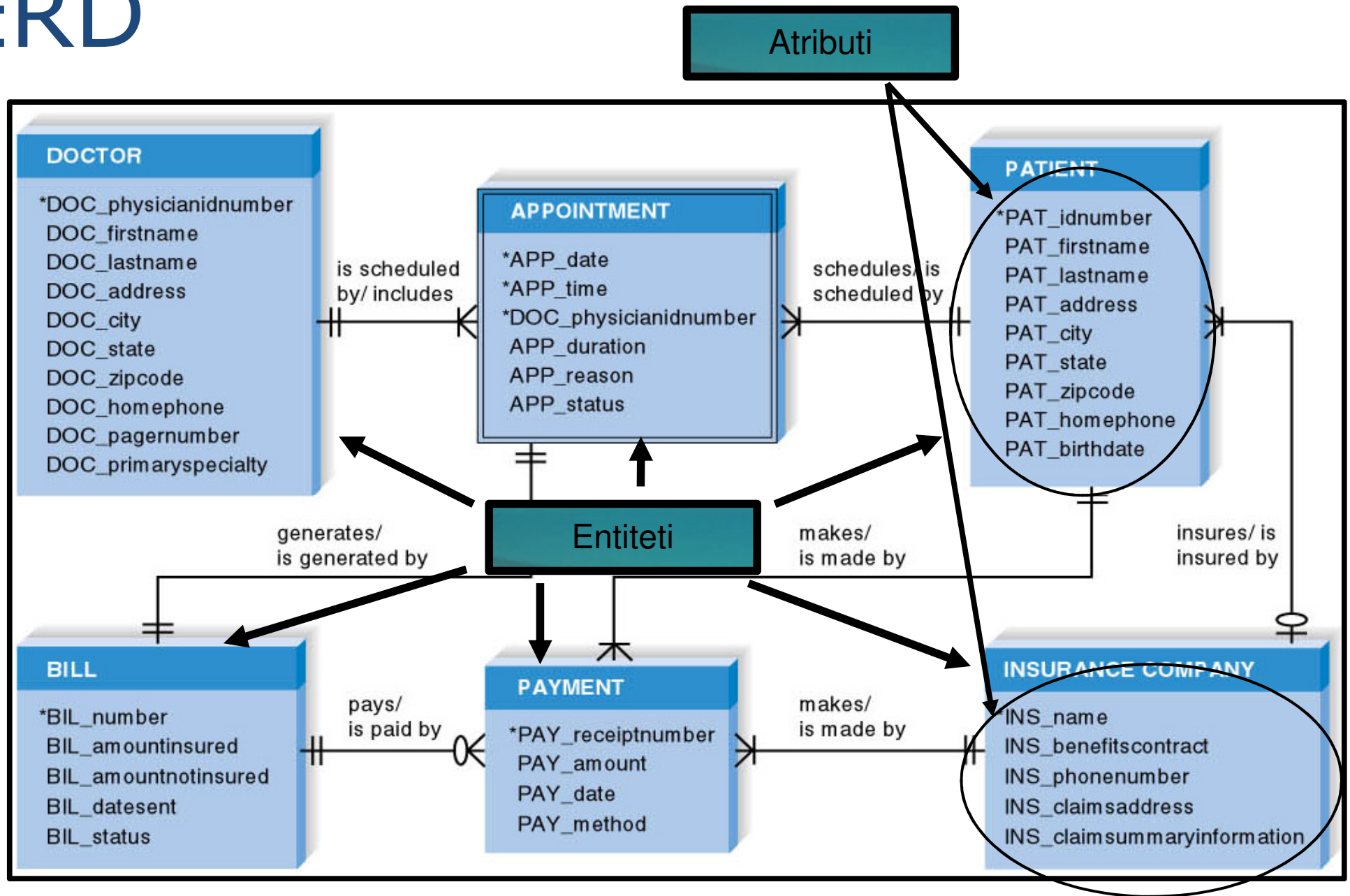
ERD

- Važno svojstvo entiteta je da predstavlja nešto što može imati puno instanci (kao klasa u OO programiranju). Npr. profesor je entitet, dok Maja Štula ne može biti entitet.
- Znači kada postoji samo jedna instanca nekog podatka onda se ne definira kao entitet.

ERD

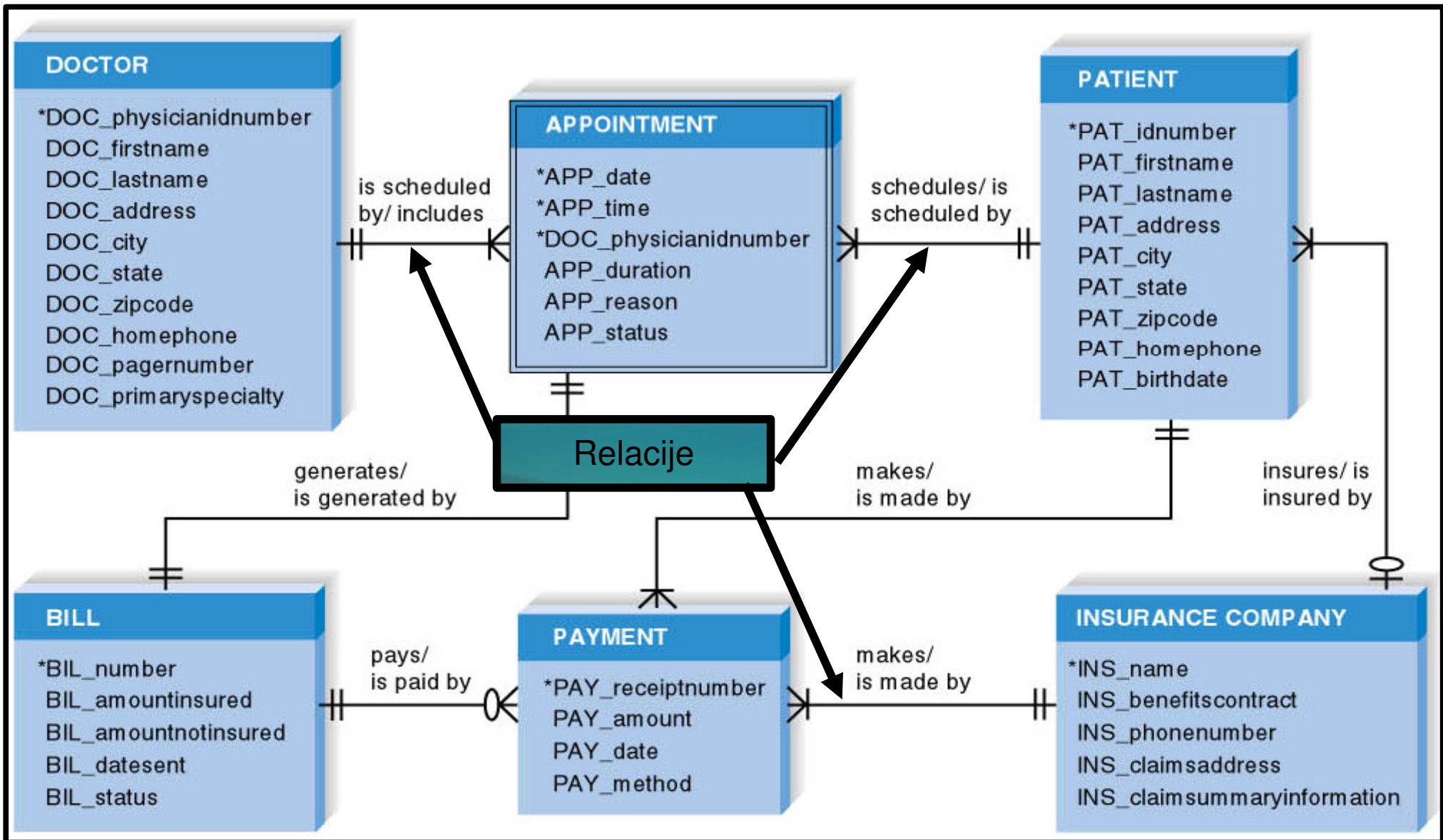
- Atributi su dostupne informacije o entitetu, označavaju se imenicom, mogu se koristiti i kao identifikatori entiteta. (npr. entitet PROFESOR ima attribute ime, prezime, smjer,).
- Za svaki entitet moguće je definirati veliki broj atributa, ali potrebno je ograničiti se samo na one attribute koji se stvarno i koriste u poslovnom procesu.

ERD



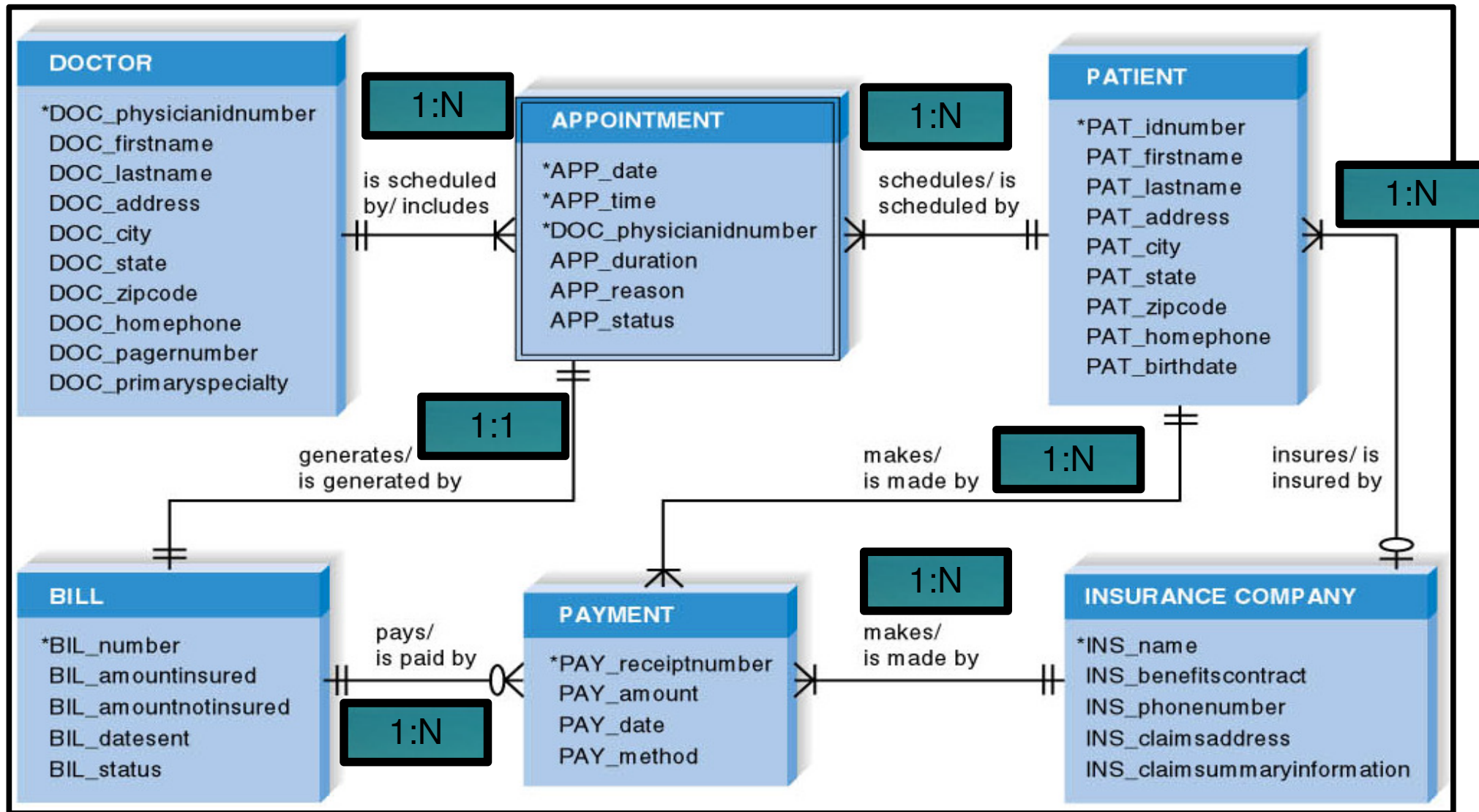
ERD

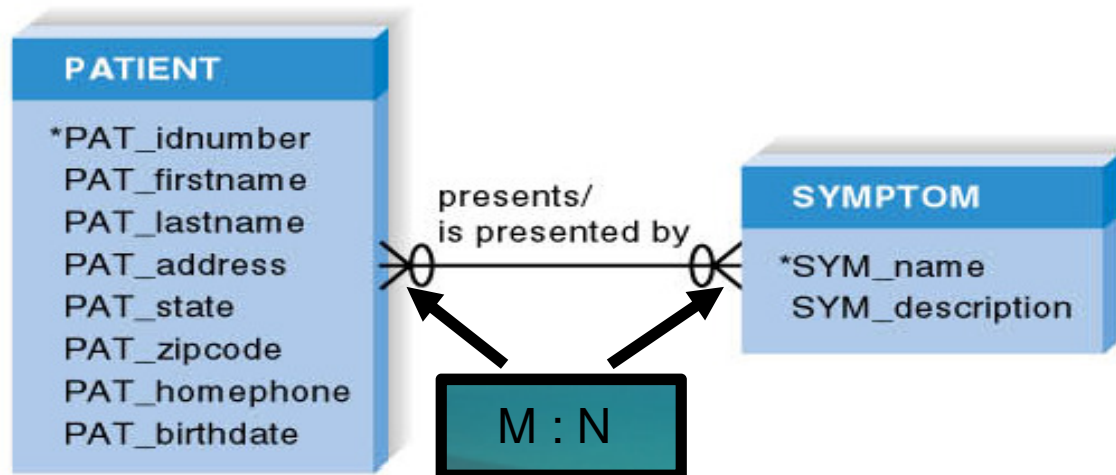
- Relacija je veza (asocijacija) među entitetima.
- Prvi entitet u relaciji je entitet roditelj, a drugi entitet je dijete.
- Relacije bi trebale biti imenovane glagolom.
- Relacije su dvosmjerne zbog toga i dva imena relacije s obzirom na smjer u kojem se relacija čita.



ERD

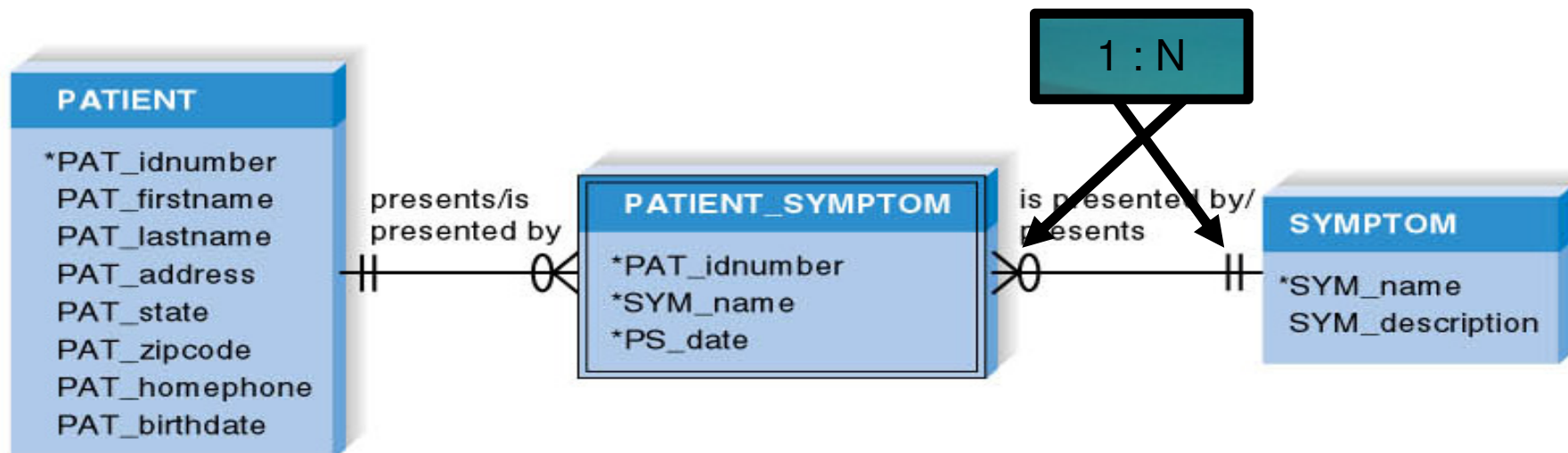
- Binarne relacije definirane su sa dva svojstva:
 - Kardinalnost - određuje koliko instanci entiteta roditelja je asocirano sa koliko instanci entiteta djeteta. Moguće kardinalnosti relacije su 1:1, 1:N ili M:N
 - Modalnost - određuje da li instanca *child* entiteta može postojati bez instance *parent* entiteta, moguće modalnosti su “nula” ili “ne nula”.

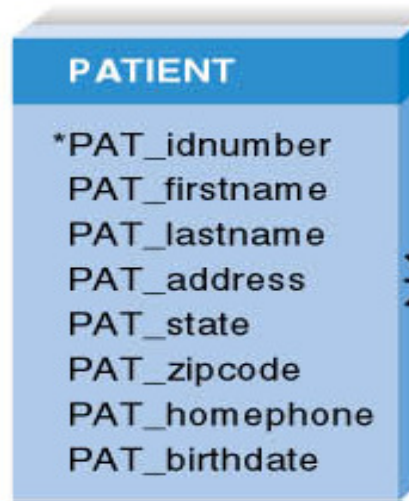




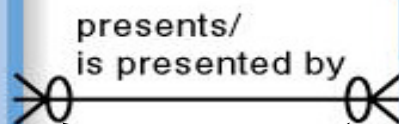
Kardinalnost 1
se označava sa |
(ravnom crtom)

Kardinalnost "više"
se označava sa tzv.
Crow's foot <



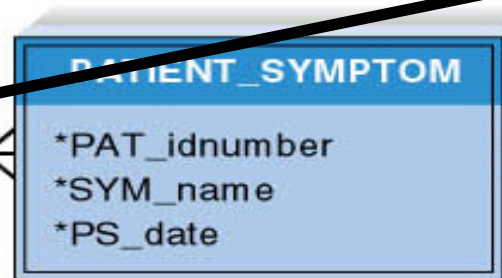
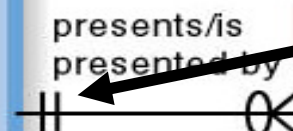
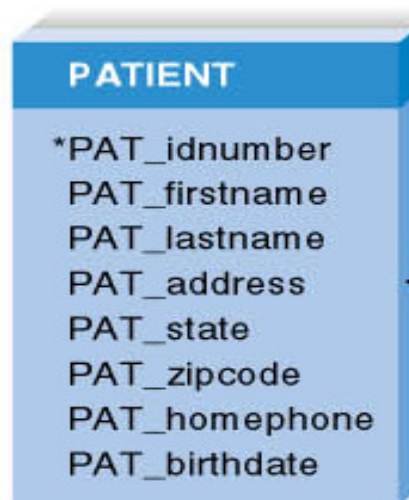


Modalnost nula
se označava sa 0

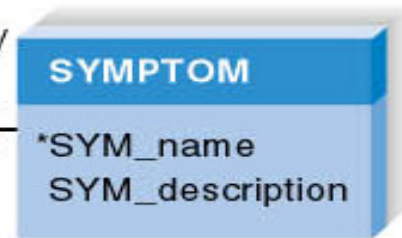
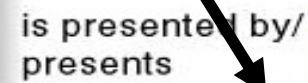


Modalnost "nula"

Modalnost "ne nula"
se označava sa 1 (ravnom
crtom)

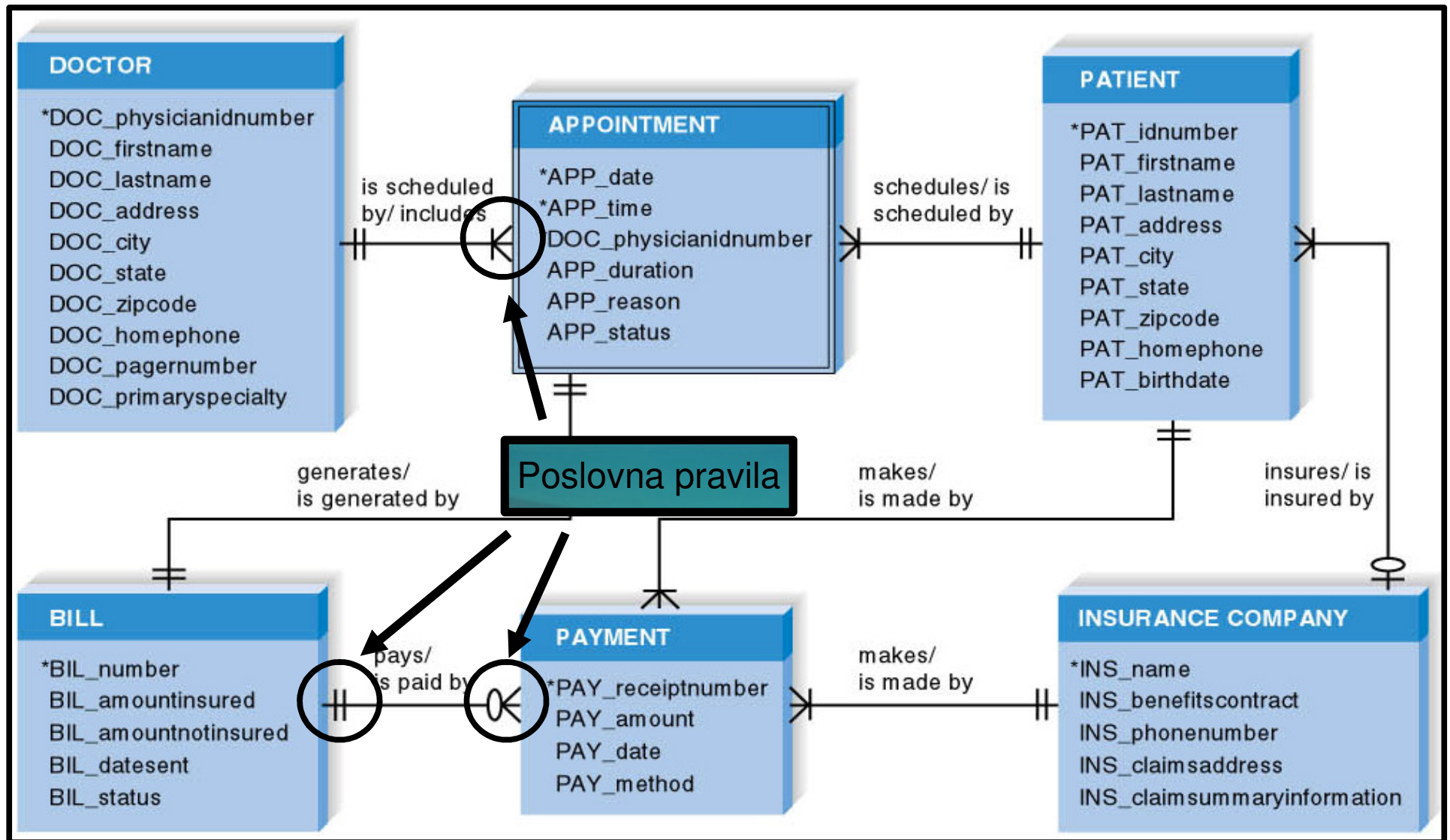


Modalnost "ne nula"



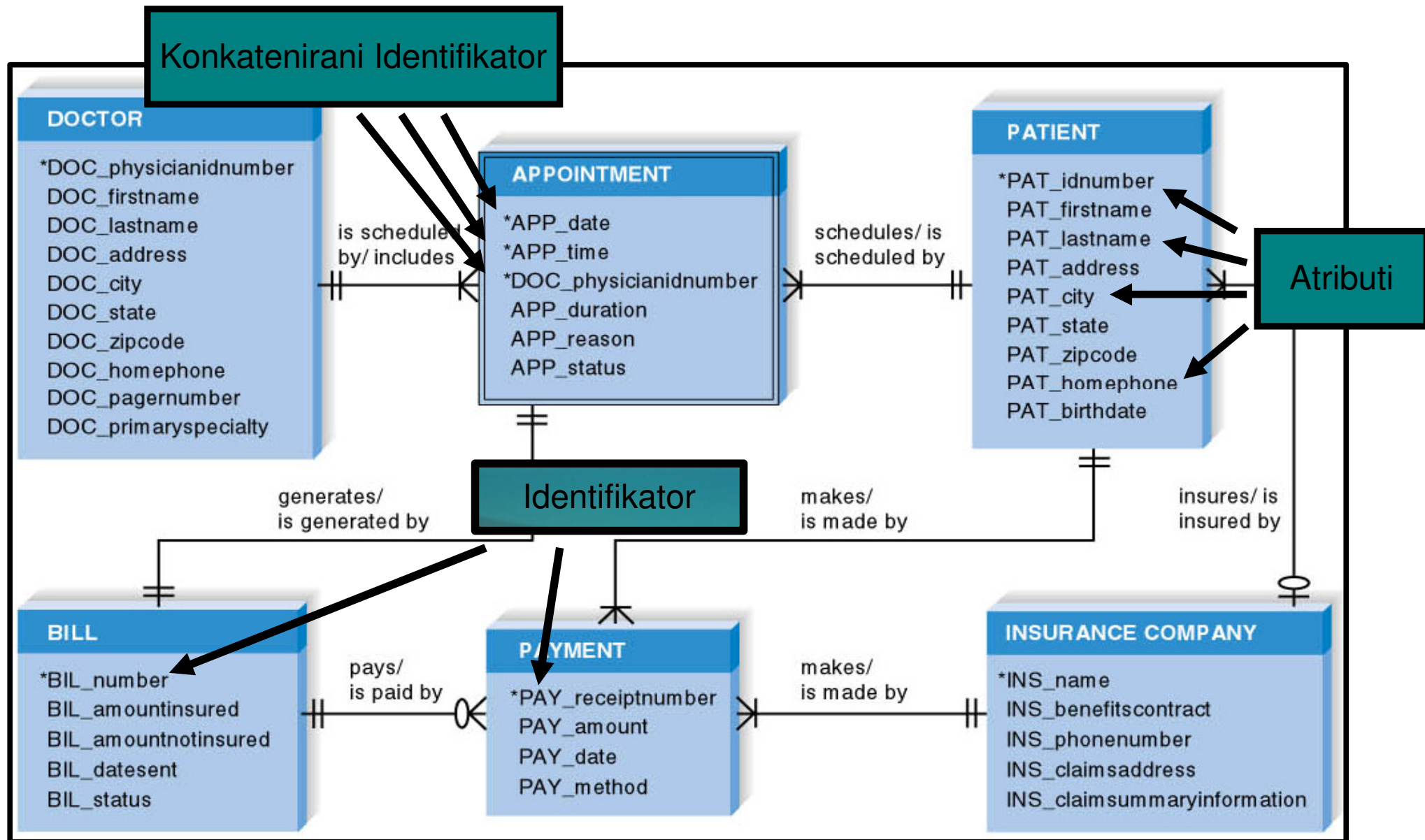
ERD

- Poslovna pravila su ograničenja koja se trebaju poštivati tijekom rada sustava.
- ERD simboli omogućavaju prikazivanje poslovnih pravila.
 - Kada instanca jednog entiteta mora postojati da bi instanca drugog entiteta postojala (npr. uplata treba imati podatke o uplatitelju – modalnost)
 - Kada je instanca jednog entiteta povezana sa samo jednom ili sa više instanci drugog entiteta (npr. jedan profesor predaje jedan ili više kolegija - kardinalnost)



Identifikatori entiteta

- Jadan ili više atributa entiteta mogu poslužiti kao identifikatori entiteta (*entity identifier*), koji jedinstveno identificiraju svaku instancu entiteta.
- Spojeni identifikator (*concatenated identifier*) se sastoji od nekoliko atributa.
- Identifikator može biti i neki “umjetni podatak” poput ID broja za određeni entitet (*ID number*).



Identifikatori entiteta

- Identifikatori se točno određuju u fazi dizajna sustava jer “umjetni podaci” (autoinkrementalni broj) ne bi trebali biti dio logičkog modela procesa, ali se češće stavljaju već u fazi analize.
- Identifikatori se na ERD dijagramu označavaju sa *.

CASE alati

- Postoji niz CASE alata za izradu podatkovnih modela.
- Neki alati (Erwin) omogućavaju izradu i održavanje logičkih i fizičkih modela podataka i automatsko generiranje baza podataka.
- Većina DBMS sustava sadrži CASE alate za razvoj modela koji se automatski prebacuju u bazu podataka koju promatrani DBMS podržava (Oracle Designer).

CASE alati

- Visible Analyst Workbench je CASE alat koji omogućava i izradu podatkovnih modela.
- Svaki CASE alat ima tzv. rječnik podataka (*data dictionary*). Rječnik podataka sadrži podatke o entitetima, atributima i relacijama ERD-a tj. sadrži metapodatke (podatke o podacima).
- Metapodaci su svi podaci koji opisuju entitet, atribut ili relaciju (ime entiteta, kardinalnost relacije, opis atributa,...).
- Rječnik podatak omogućava djeljenja metapodataka između suradnika na projektu i pruža jedinstveno, centralno mjesto pohrane i održavanja tih podataka.

Investment Research Database Version 1.02 - as at 1996 03 17 17:16

Copyright © 1996 3QC Funds

Contact Table**Contact Table**

A business or other contact.

Created on the workstation, replicated to the server, read only on the server.

Unload/load data: t2contac

End users have select, insert and update rights.

Comments char(1000) not null default "

Free-form comments about this contact.

Contact_id integer not null default AutoIncrement (PK)

Automatically generated identifier, unique on the workstation database.

Contact_subscription_id serves to differentiate between contacts on the server.

Contact_subscription_id char(6) not null (PK) (FK)

Subscription identifier for publish subscribe replication.

Date_time_updated timestamp not null default timestamp

The date and time this contact was first added or last changed.

Fax char(20) not null default "

Fax number of the contact.

First_name char(50) not null default "

First name of the contact.

Home_page char(100) not null default "

World Wide Web URL.

ERD

- Entiteti se dijele na tri tipa:
 1. Zavisni entitet:
 - u slučaju kada dijete entitet zahtjeva attribute od entiteta roditelja da bi se jedinstveno identificirala instanca dijeteta onda se dijete entitet naziva zavisni entitet
 - relacija koja uključuje zavisni dijete entitet naziva se identificirajuća relacija
 2. Nezavisni entitet:
 - entitet koji se može identificirati bez pomoći nekog drugog entiteta
 - relacija koja uključuje nezavisni dijete entitet zove se neidentificirajuća relacija

ERD

3. Presjecajući entitet:

- postoji u svrhu prikupljanja informacija o relaciji između dva entiteta
- dodaju se u podatkovni model da bi pohranili informacije o dva entiteta koji dijele M:N relaciju
- nazivaju se još i asocijativni entiteti

Presjecajući entitet

- Proces dodavanja presjecajućeg entiteta naziva se razdvajanje M:N relacije.
- Presjecajući entitet se dodaje u tri koraka:
 1. uklanjanje M:N relacije i umetanje novog entiteta između dva postojeća
 2. dodavanje dvije 1:N relacije u podatkovni model
 3. dodjela imena presjecajućem entitetu

Postupak izrade ERD

- Postupak izrade ERD-a se sastoji od tri koraka:
 1. Identifikacija entiteta
 2. Identifikacija atributa entiteta
 3. Identifikacija relacija

Postupak izrade ERD

- Izrada kreće ili od modela procesa ili ako model procesa nije postavljen od identificiranih zahtjeva i slučajeva korištenja.
- Spremišta podataka i tokovi podataka iz modela procesa služe za identifikaciju entiteta i atributa.
- Model podataka treba biti balansiran sa modelom procesa.

Postupak izrade ERD

- Postupak izrade ERD je iterativan.
- Ako se u izradu ERD-a kreće od slučajeve korištenja onda se uzimaju podaci o ulazima i izlazima slučajeve korištenja.
- Ukoliko se radi o složenom ERD-u, može ga se podijeliti na više zasebnih dijagrama grupiranih prema sličnosti entiteta.

Validacija ERD

- Validacija ERD se provodi u nekoliko koraka.
- Prvo se provodi balansiranje modela procesa i podatkovnog modela.
- Nakon toga se postupkom normalizacije dobiva konačni oblik podatkovnog modela.

Balansiranje ERD-a i DFD-a

- Iako se model procesa fokusira na procese poslovnih sustava, sastoji se od dvije podatkovne komponente:
 - tok podataka
 - spremište podataka
- Svrha je da se ilustrira koji podaci su korišteni i kreirani od strane procesa i gdje se drže, tj. gdje su pohranjeni ti podaci.
- Te komponente dijagrama toka podataka trebaju biti balansirane s dijagramom entiteta i relacija.
- Mnogi CASE alati nude mogućnost identifikacije problema kod balansiranja između DFD-a i ERD-a.

Balansiranje ERD-a i DFD-a

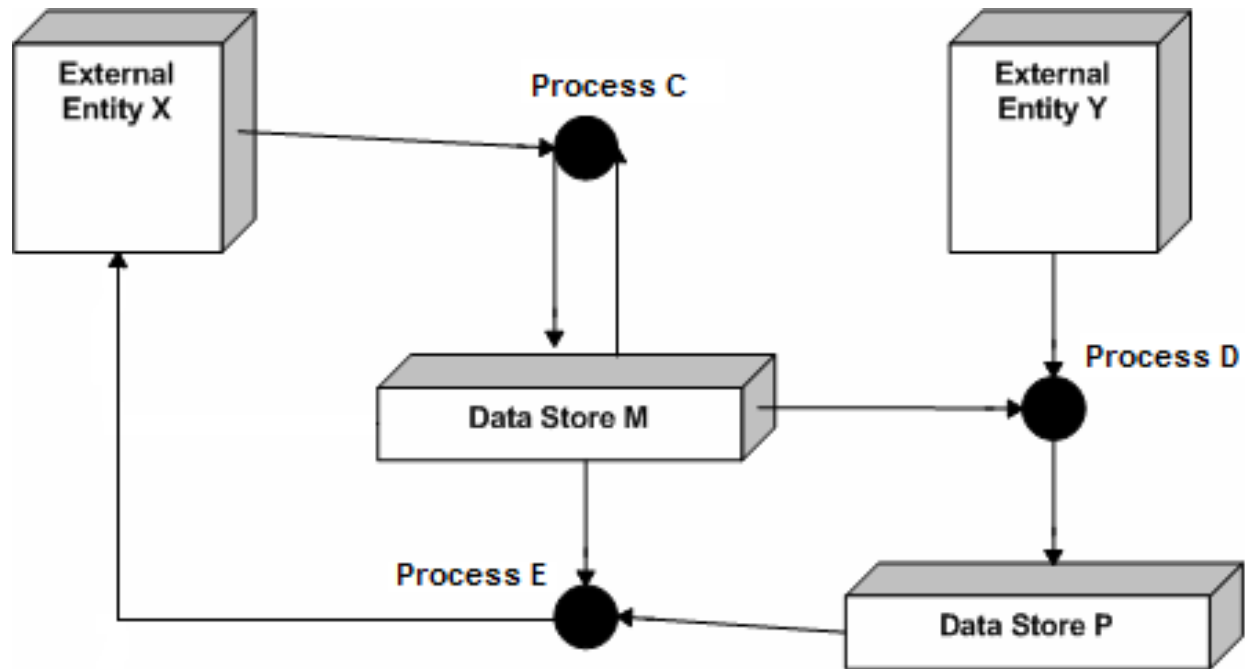
- Trebamo provjeriti da li su svi podaci u spremištima i tokovima podataka u procesnom modelu uključeni negdje kao entitet ili atribut entiteta u modelu podataka.
- Želimo osigurati da model u potpunosti sadrži podatke predstavljene u procesnom modelu.
- Ako to nije osigurano onda je model podataka nepotpun.
- Svi podatkovni elementi trebali bi se pojaviti kao dio spremišta podataka i toka podataka u procesnom modelu.
- Ako neki podatkovni elementi nedostaju u procesnom modelu, trebamo provjeriti da li su ti podaci zaista potrebni u radu sustava.

CRUD

- Koristan alat za jasno prikazivanje relacija između procesnog modela i modela podataka je CRUD (Create, Read, Update, Delete) matrica.
- To je tablica koja opisuje kako procesi koriste podatke unutar sustava.
- CRUD matrica(e) se najprije razvija na logičkom procesu i modelu podataka, a onda se preispituje u fazi dizajna.
- Matrica također pruža važne informacije za specifikaciju programa zato jer pokazuje točno kako su podaci korišteni i stvarani od strane glavnih procesa sustava.

Kreiranje CRUD matrice

- Da bi kreirali CRUD matricu, nacrtamo tablicu navodeći sve procese u prvoj liniji, a entitete i njihove attribute u prvom stupcu s lijeve strane (ili obrnuto).
- Zatim ovisno o informacijama u procesnom modelu analitičar ispunjava svaku ćeliju tablice sa C,R,U,D slovom ili ostavlja praznu da bi opisao interakciju pojedinog procesa sa pojedinim entitetom.



	Process C	Process D	Process E
Data Entity M			
Attribute M-1	CRUD	R	R
Attribute M-2	CRUD		R
Attribute M-3	CRUD	R	
Attribute M-4	CRUD		R
Data Entity P			
Attribute P-1		C	R
Attribute P-2		C	
Attribute P-3		C	R

CRUD matrica

- Možemo identificirati mjesta gdje su atributi možda izostavljeni iz spremišta podataka ili entiteta.
- Nadalje, možemo provjeriti da se svaki atribut kreira, čita, ažurira i briše negdje u procesnom modelu.
- Ako nije pročitano od strane nekog procesa, onda atribut vjerojatno nije potreban.
- Ako nije kreiran ili ažuriran, atribut vjerojatno treba dodati u tok podataka u procesnom modelu.

Validacija ERD-a

- Normalizacija je tehnika koja se koristi za validaciju podatkovnih modela.
- To je proces u kojem se niz pravila primjenjuje na logički model da bi se odredilo koliko je dobro model formiran.
- Uobičajeno se koriste tri normalizacijska pravila ili forme.
- Normalizacijom se iz postojećih podataka nastoji ukloniti redundantne podatke tj. iste podatke koji postoje na dva mjesta (direktna redudancija) ili podatke koji se mogu izvući iz drugih podataka (indirektna redudancija), a zadržati integritete podataka.

Normalizacija – 1NF

- Logički model podataka je u prvoj normalnoj formi (1 Normal Form (1NF)) ako ne sadrži ponavljajuće atribute (atributi koji sadrže višestruke vrijednosti za jednu instancu).
- Podaci se dovode u 1NF:
 1. Uklonjanjem višestrukog pojavljivanja istog stupca u tablici.
 2. Kreiranjem posebne tablice za svaki skup povezanih podataka te identificiranjem jedinstvenim ključem svakog redka tablice.

Normalizacija – 1NF

- Što ako dođe treći student?
- Što ako se promjeni prezime jednog studenta?

Profesor	Student 1	Student 2	Student 3
Štula Maja	Dodig Branko	Miletić Zlatko	Tomaš Srećko
.	.	.	.
.	.	.	.
.	.	.	.



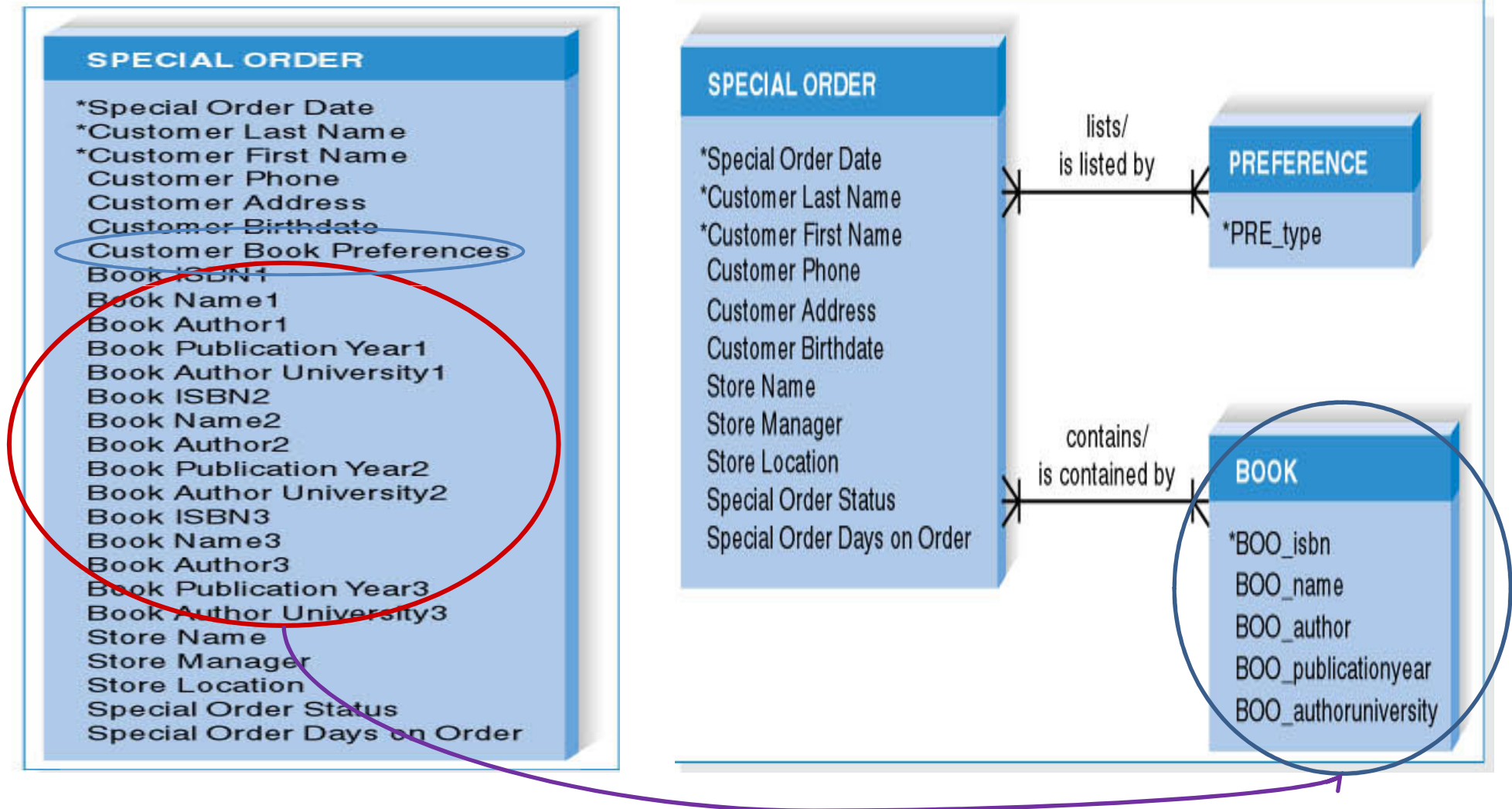
Rješenje

ID_profesor	Profesor
1	Štula Maja
.	.
.	.
.	.

ID_profesor	ID_student
1	1
1	2
1	3
.	.
.	.
.	.

ID_student	Student
1	Dodig Branko
2	Miletić Zlatko
3	Tomaš Srećko
.	.
.	.
.	.

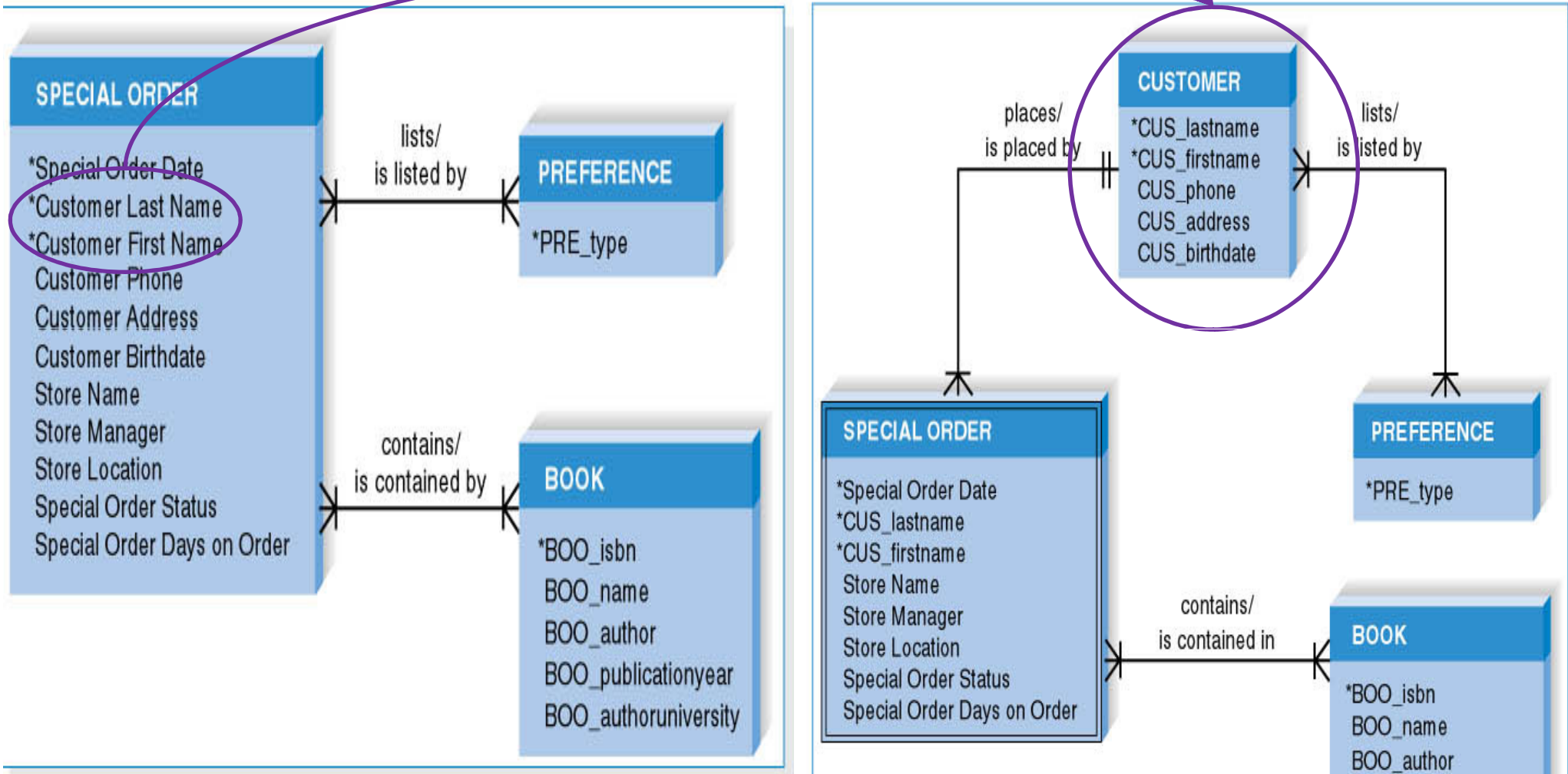
Normalizacija – 1NF



Normalizacija – 2NF

- Druga normalna forma (2NF) zahtjeva da svi entiteti budu u prvoj normalnoj formi (1NF).
- Podaci se dovode u 2NF:
 1. Ako entitet ima spojeni identifikator potražiti attribute koji ovise samo o dijelu identifikatora i odvojiti ih u zasebni entitet.

Normalizacija – 2NF



Normalizacija – 3NF

- Treća normalna forma (3NF) se može postići kada je model već i u prvoj (1NF) i u drugoj (2NF) normalnoj formi.
- Podaci se dovode u 3NF:
 1. Trebaju biti zadovoljeni svi uvjeti za prvu i drugu normalnu formu.
 2. Potražiti attribute koji ovise samo o drugom atributu koji nije identifikator i prebaciti ih u novi entitet tj. iz tablice treba ukloniti kolone koje nisu u potpunosti ovisne o primarnom ključu.

Normalizacija

