

Database Administration:

Lekcija 3

Modelovanje Podataka & Normalizacija



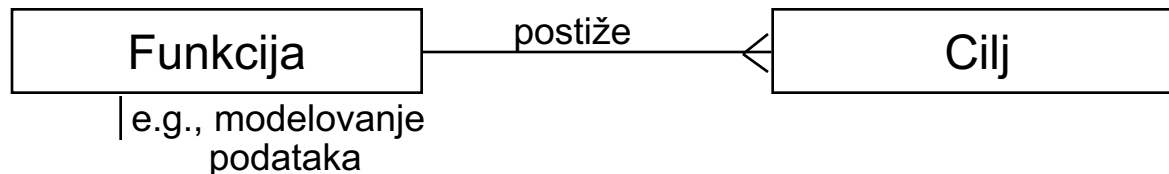
Sadržaj lekcije

- Koncepti modelovanja podataka
- Komponente modela podataka
- Pronalaženje entiteta, atributa i veza
- Konceptualni, Logički i Fizički modeli podataka
- Normalizacija
- Dodatni problemi u modelovanju podataka

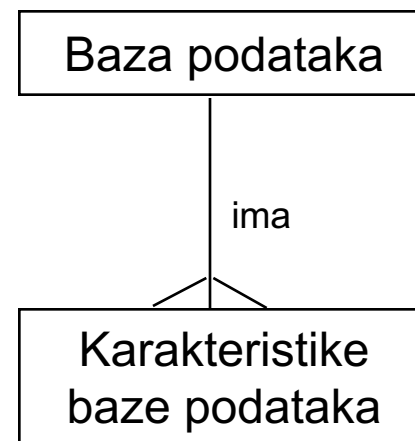
Razmišljanje o podacima

- Ne razmišljati “kako”; razmišljati “šta”
- Ne razmišljati o fizičkom; razmišljati konceptualno
- Ne razmišljati o procesu; razmišljati o strukturi
- Ne razmišljati o navigaciji; razmišljati o vezi

Ciljevi modelovanja podataka



- Dokumentovati i informisati se o zahtevima za podacima u poslovanju
- Omogućiti bazu podataka sa:
 - Minimalnom redundansom podataka
 - Maksimumom integriteta podataka
 - Boljom deljivošću podataka
 - Povećanom stabilnošću
 - Fleksibilnost
 - Povećanom konzistencijom
 - Pristupom podacima na vreme
 - Korisnošću podataka
- Povećati vrednost resursa podataka



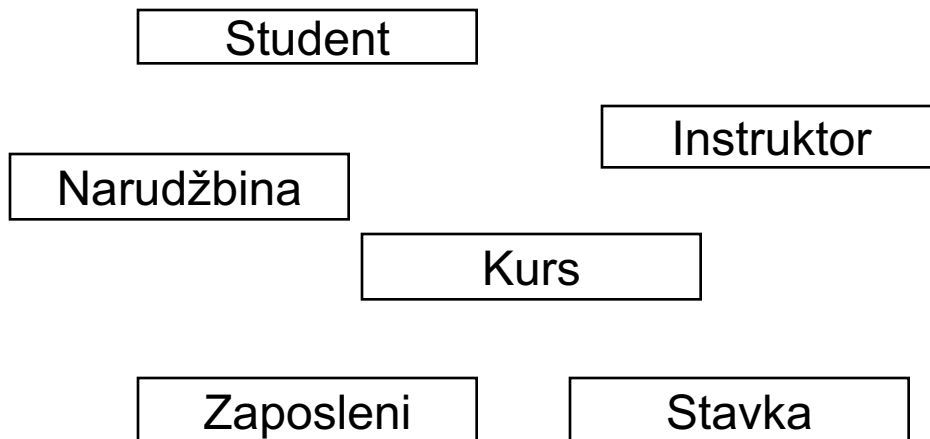
Komponente modela podataka

Modeli podataka se sastoje iz više komponenti:

- Entiteti
- Tipovi Entiteta
- Instance Entiteta
- Veze
- Atributi
- Uloge Atributa
- Ključevi
- Tehnike Dijagrama

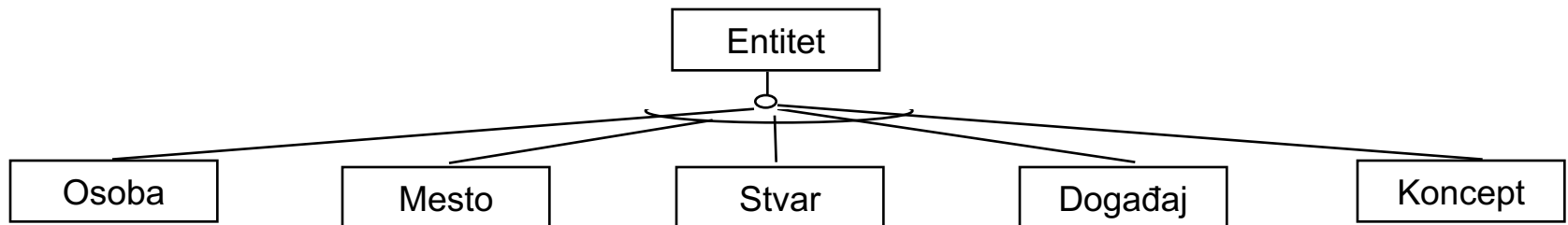
Entiteti

- Nešto što postoji i može se opisati.
- Osoba, mesto, stvar, koncept ili događaj o kojima organizacija čuva podatke



Primeri entiteta

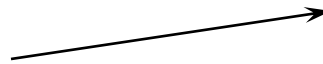
- **Osobe:** Uloge osoba
 - Zaposleni
 - Dobavljač
 - Klijent
 - Agent
- **Stvari:** Fizička vrsta objekta
 - Stavka
 - Proizvod
 - Materijal
 - Deo
 - Zgrada
 - Oprema
- **Koncepti:** Ideje
 - Upozorenje
 - Račun
 - Ruta
- **Mesta:** Područja geografske lokacije
 - Kancelarija
 - Grad
 - Skladište
 - Region
- **Događaj:**
 - Prodaja
 - Transfer
 - Projekat
 - Rezervacija
 - Narudžbina
 - Isporuka
 - Dogovor
 - Let



Instance entiteta



Tip Entiteta



Customer: Moore
City: Pittsbu
State: PA
Phone: 555
Credit:: Go
AMEX:
VISA:
Contact:

Customer: Jackson
City: An
State: MI
Phone: 5
Credit:: E
AMEX:
VISA:
Contact:

Customer: Mullins
City: Houston
State: TX
Phone: 555-1234
Credit:: Fair
AMEX: xxxxxxxxxxxxxxxxxxxxxx
VISA: xxxxxxxxxxxxxxxxxxxxxx
Contact: Mike

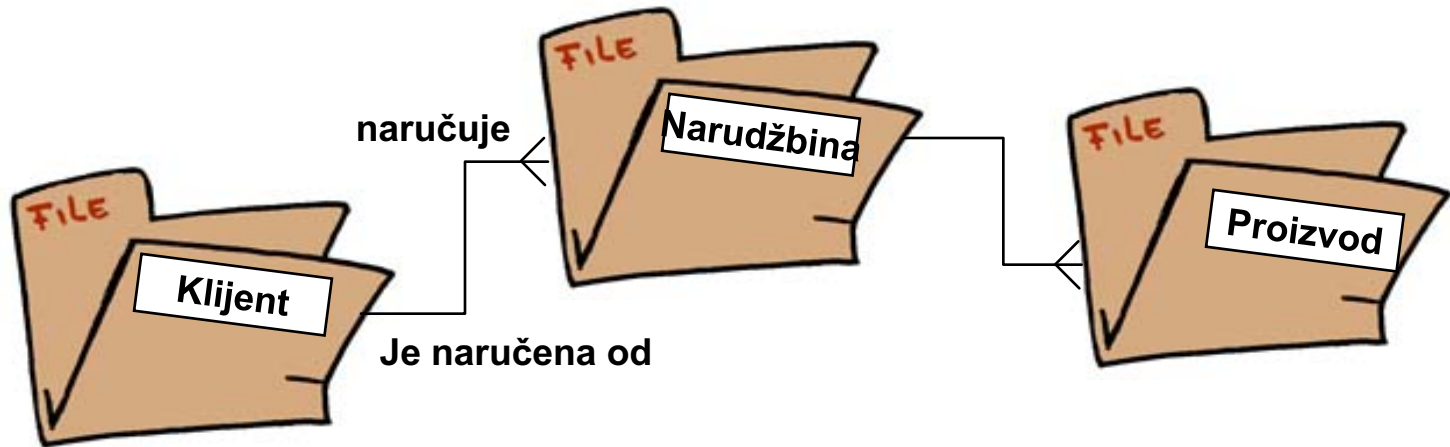


**Instance
Entiteta**

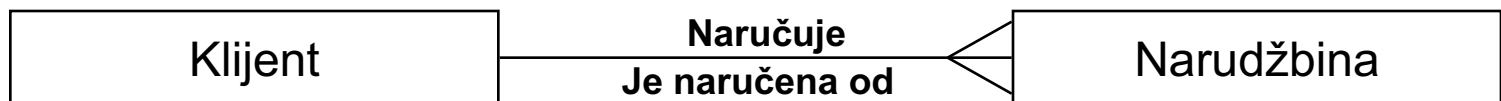
Smernice za imenovanje entiteta

- Imenica, ili format imenice sa pridevom
 - Ugovor
 - LizingUgovor
- Minimizirati korišćenje prideva
 - Ugovarač vs UgovorZaposlenog
 - Trening vs KursZaposlenog
- Ne koristiti množine; koristiti jedninu: Osoba vs Osobe
 - Tip entiteta je model ili **šablon** pre nego **skup** svih osoba
- Koristiti termine poslovanja
 - Vendor ili Supplier?
 - Korisnik ili Klijent?
- Budite konzistentni
- Uklonite artefakte iz imena entiteta koji su vezani za procese
 - Država **ne** DržavaStanovanja
 - Agent **ne** AgentProdaje

Veze



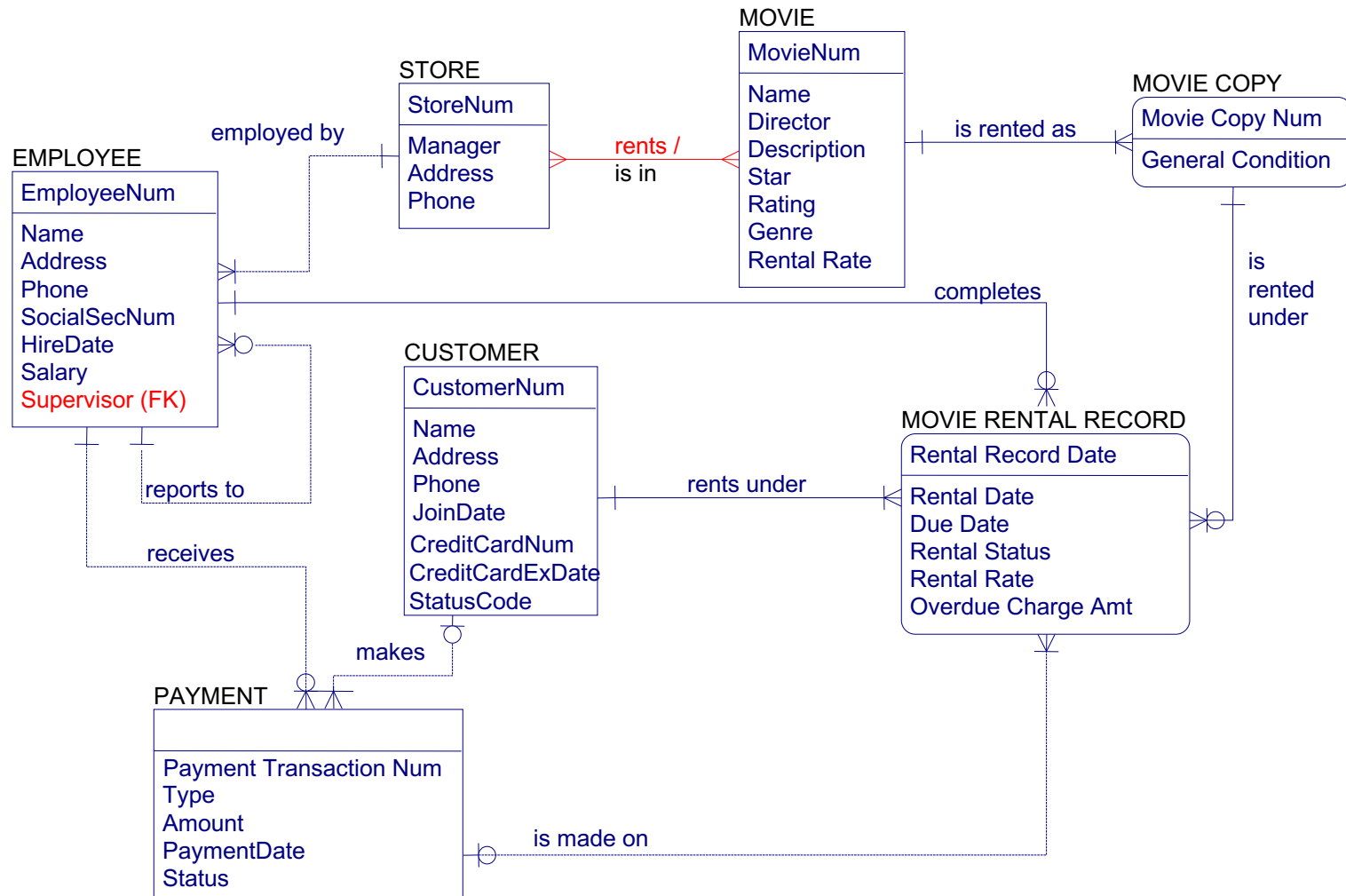
- Kako su različiti entiteti povezani jedan sa drugim



Kardinalnost



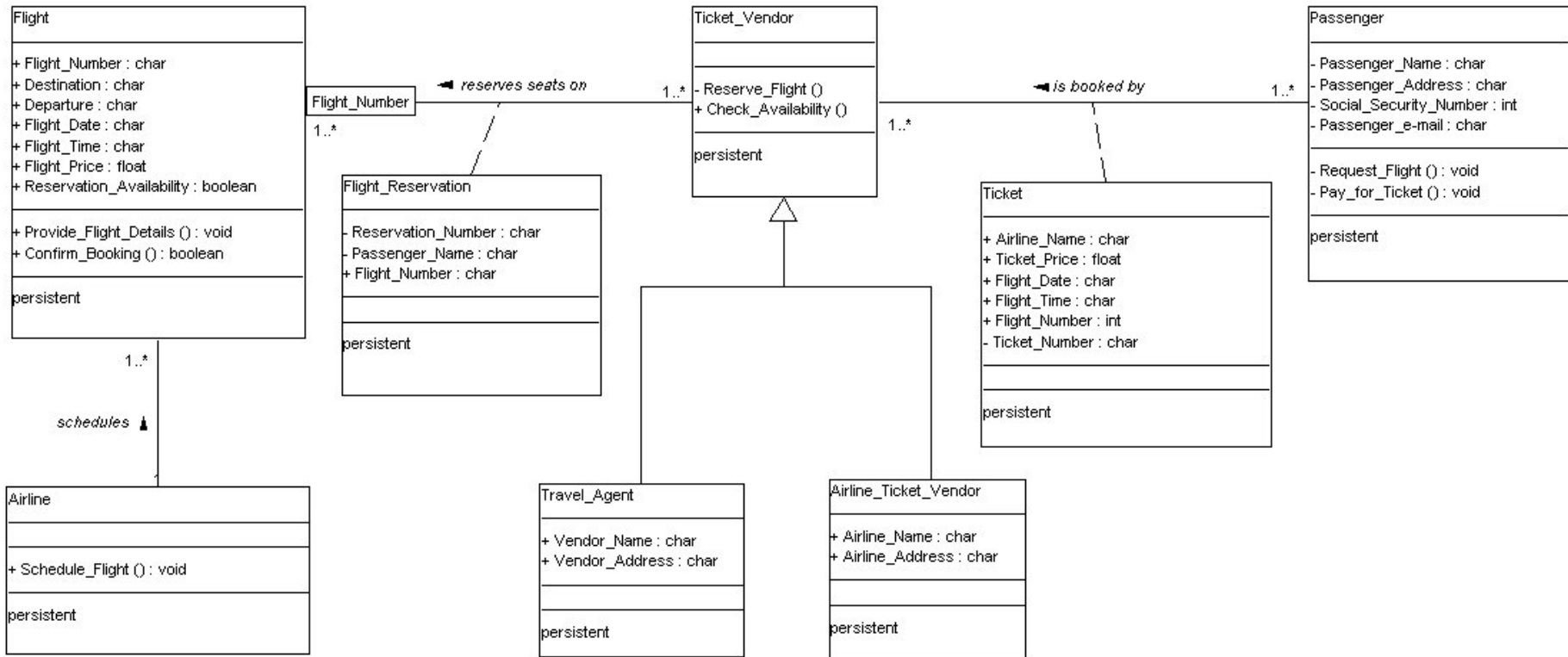
Dijagram Entiteti-Veze



Konvencije za crtanje dijagrama

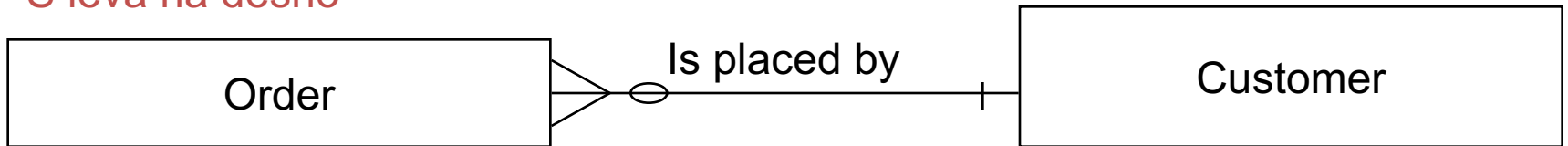
	jedan-prema-jedan	jedan-prema-više	više-prema-više
1			
2			
3			
4			
5			

UML

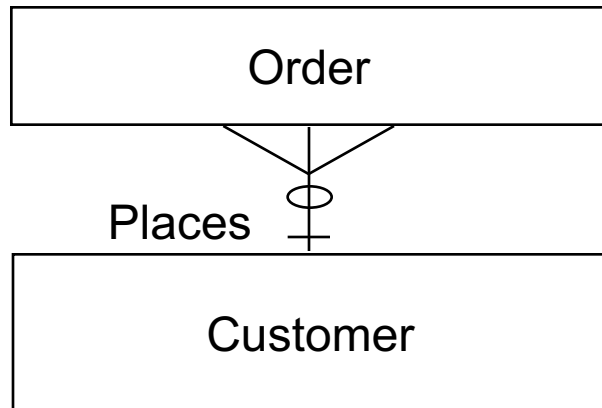


Smernice za modelovanje dijagrama

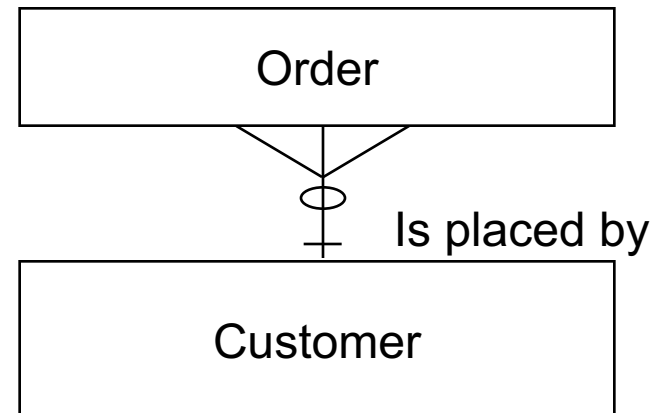
S leva na desno



Gore ka dole



Dole ka gore



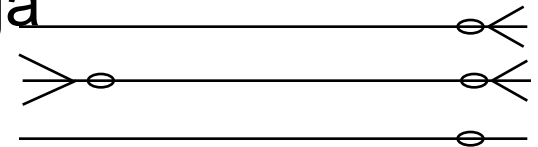
s desna na levo



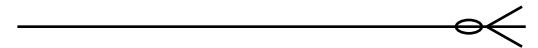
Tipovi Entiteta

- Analiza veza može da nam pomogne da odredimo tip entiteta:

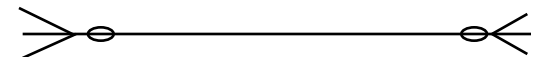
- **Osnovni Entiteti** – predstavljaju osnovne, ili fundamentalne objekte poslovanja



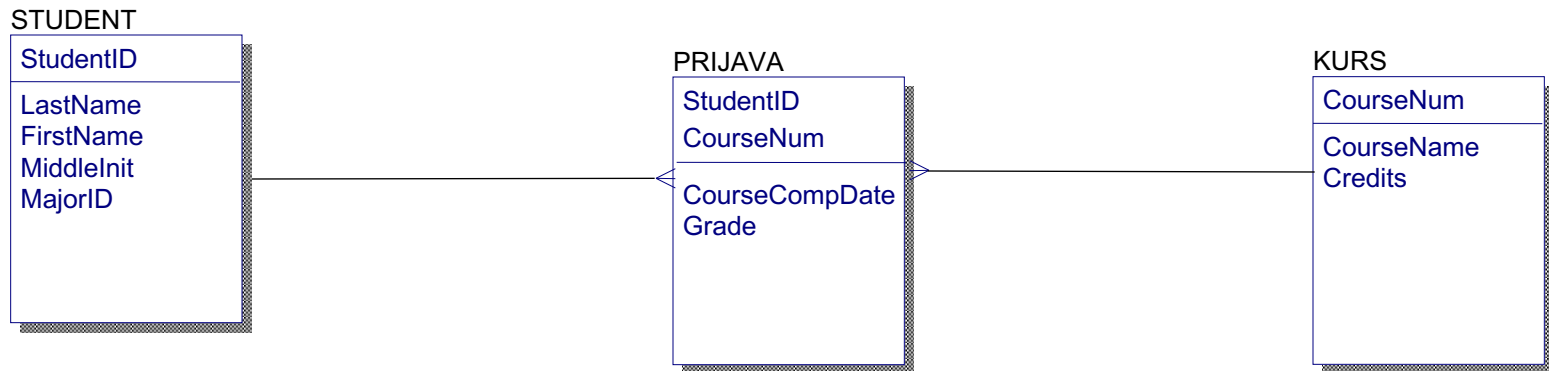
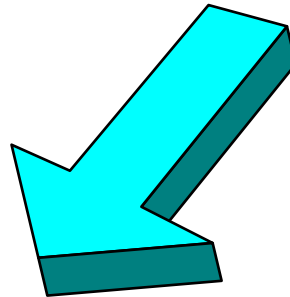
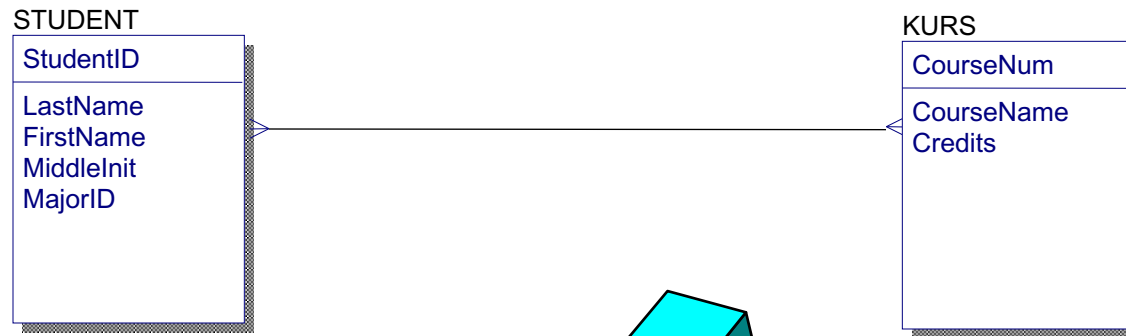
- **Entiteti Karakteristika** – sadrže više atributa ili činjenica koji opisuju osnovni entitet



- **Među Entiteti** – opisuje vezu između dva druga entiteta



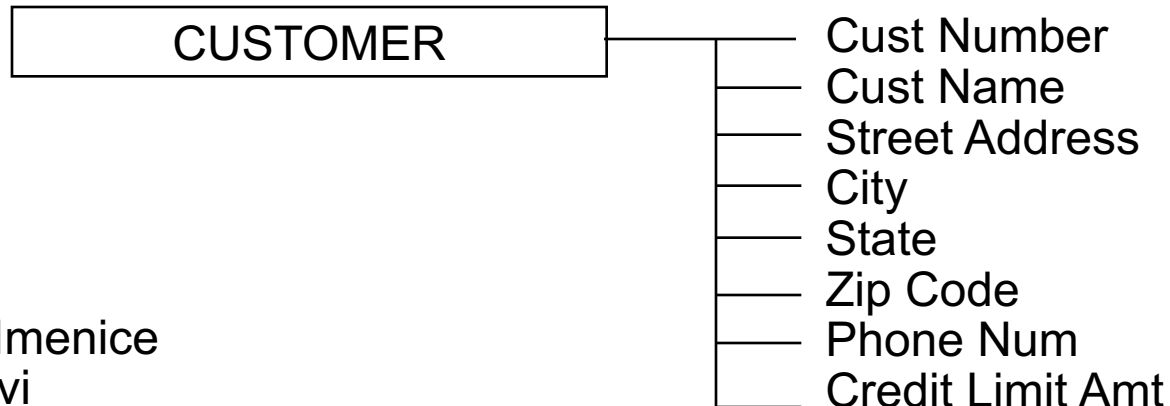
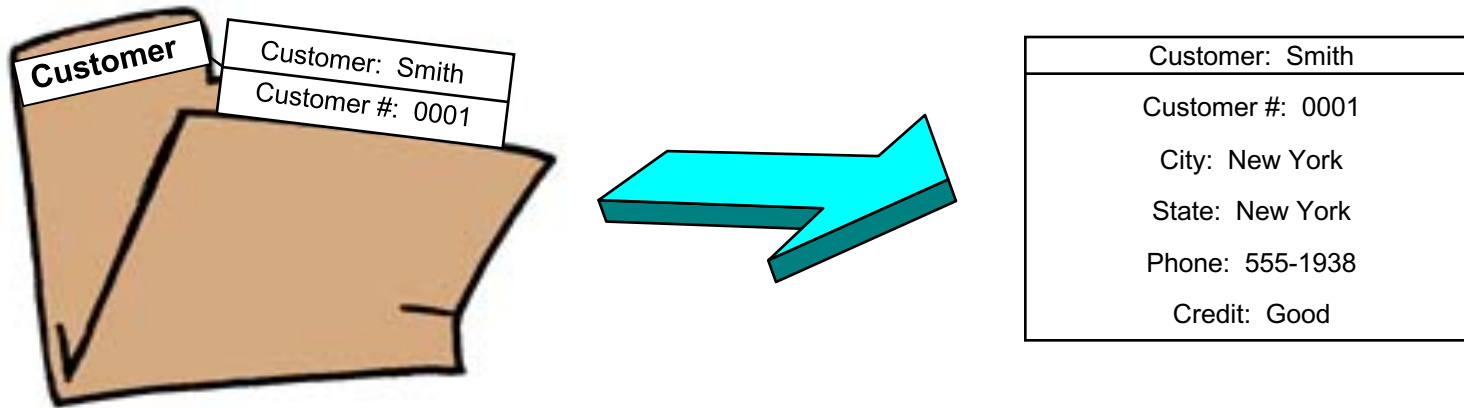
Kreiranje Među-Entiteta



Atributi

- ♦ Atribut je činjenica o entitetu.
- ♦ Svaki atribut mora da ispuni jedan od sledećih ciljeva:
 - Identifikuje
 - Povezuje
 - Opisuje

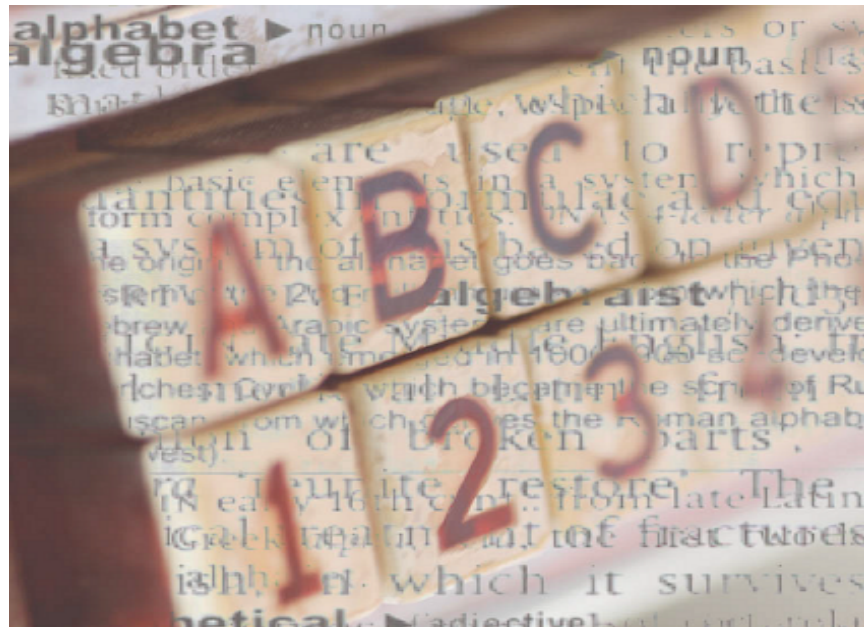
Atributi



ENTITET = Imenice
ATR = Pridevi

Domeni

- Domen definiše dozvoljene vrednosti za atribut
- Svakom atributu je dodeljen domen



Smernice za imenovanje atributa

Klasa	Skraćenica	Opis
ADDRESS	ADDR	Adresa
AMOUNT	AMT	Monetary amount
CODE	CODE	Classifications, codes, types
DATE	DATE	Calendar dates
DESCRIPTION	DESC	Descriptive text
IDENTIFIER	ID	Alphanumeric unique identifier
IMAGE	IMG	Encoded digital image
NAME	NAME	Alphabetic name, identification
NUMBER	NUM	Numeric count
PERCENT	PCT	Relationships as %
QUANTITY	QTY	Counts and units
TEXT	TXT	Free form document text
TIME	TIME	Time; temporal data

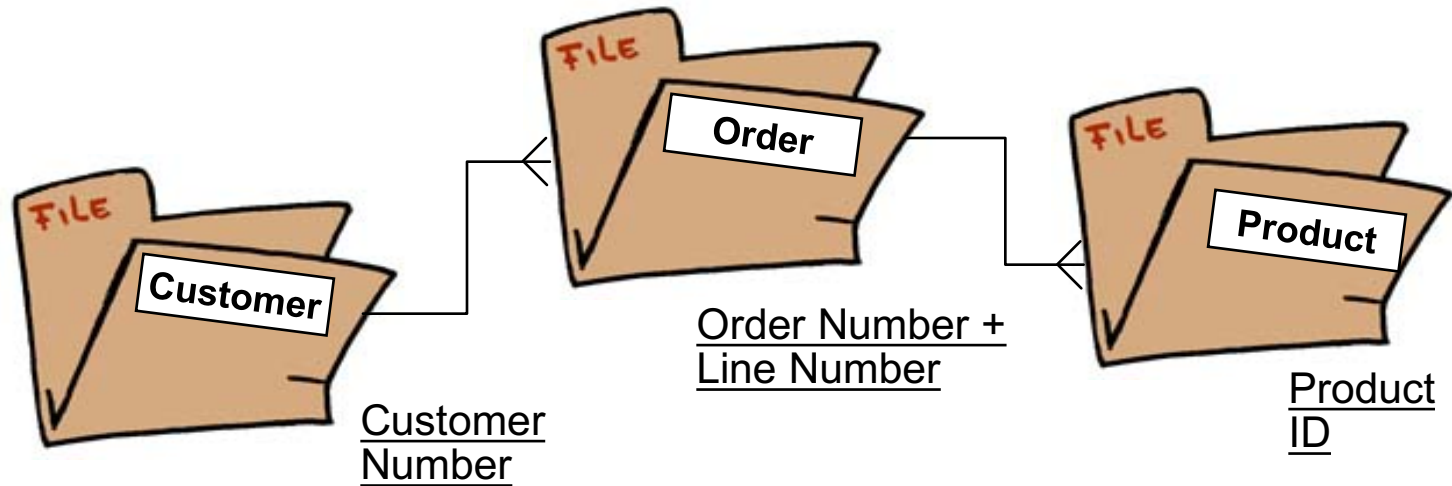
Vrednosti Atributa i NULL

- Atributi i vrednosti atributa su slični entitetima i instancama
- Vrednost atributa je instanca atributa
- Ako vrednost atributa može da bude null to znači jednu od dve stvari:
 - atribut nije primenljiv kod nekih instanci atributa, ili
 - atribut je primenljiv za sve instance, ali informacije mogu da budu nepoznate.

Ključevi

- ♦ Ključ je atribut ili grupa atributa koji se koriste za identifikaciju
- ♦ Postoji tri vrste ključeva:
 - Ključevi kandidati
 - Primarni Ključevi
 - Strani ključevi
- ♦ Ključ ne bi trebalo da sadrži skriveno značenje
 - Svrha ključa jeste da identifikuje, ne da opisuje

Primarni Ključevi



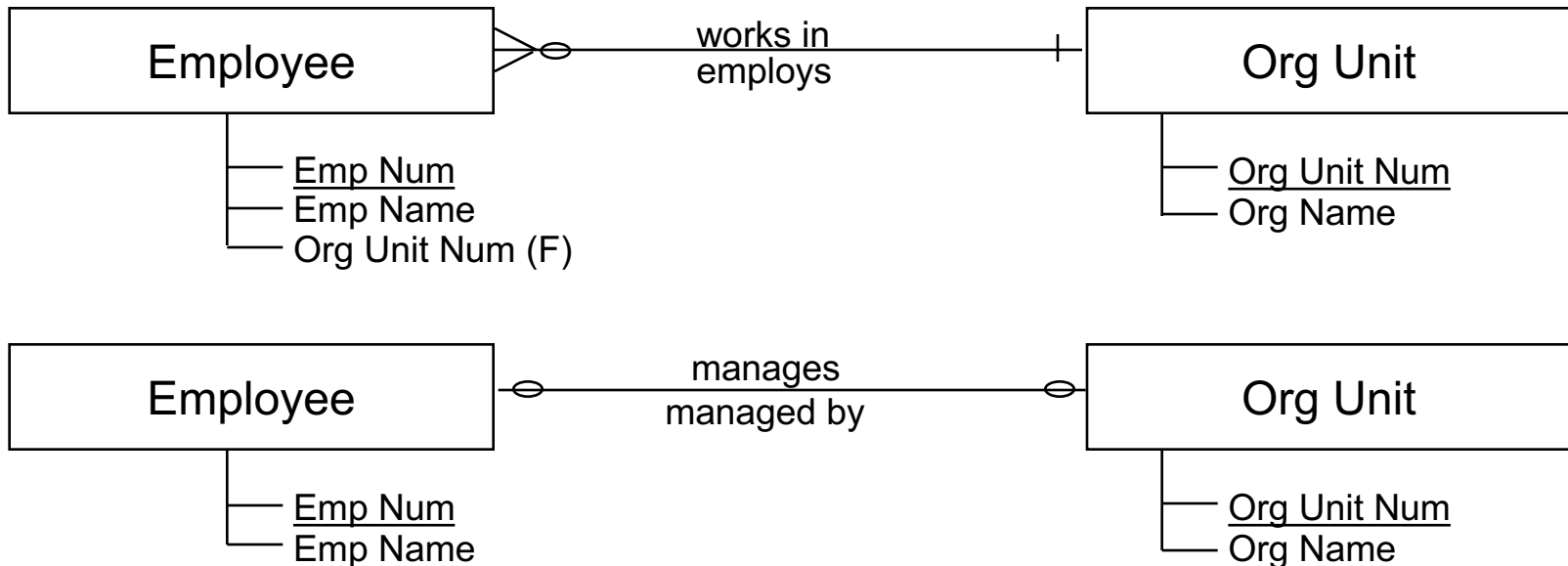
- Atribut ili atributi koji jedinstveno identifikuju entitet
- Vrednost primarnog ključa jedinstveno identifikuje jednu instancu entiteta



Strani Ključevi

(Atributi Veze)

- Atribut(i) u jednom entitetu koji povezuju instancu jednog entiteta sa instancom drugog entiteta
- Povezujući atribut(i) sadrže iste vrednosti kao primarni ključ odgovarajućeg entiteta



Ključevi Kandidati

- Može da postoji nekoliko mogućih primarnih ključeva
 - Moraju da identifikuju tu instancu entiteta
- Oni su poznati kao ključevi kandidati
- Trebalo bi da postoji jedan i samo jedan primarni ključ



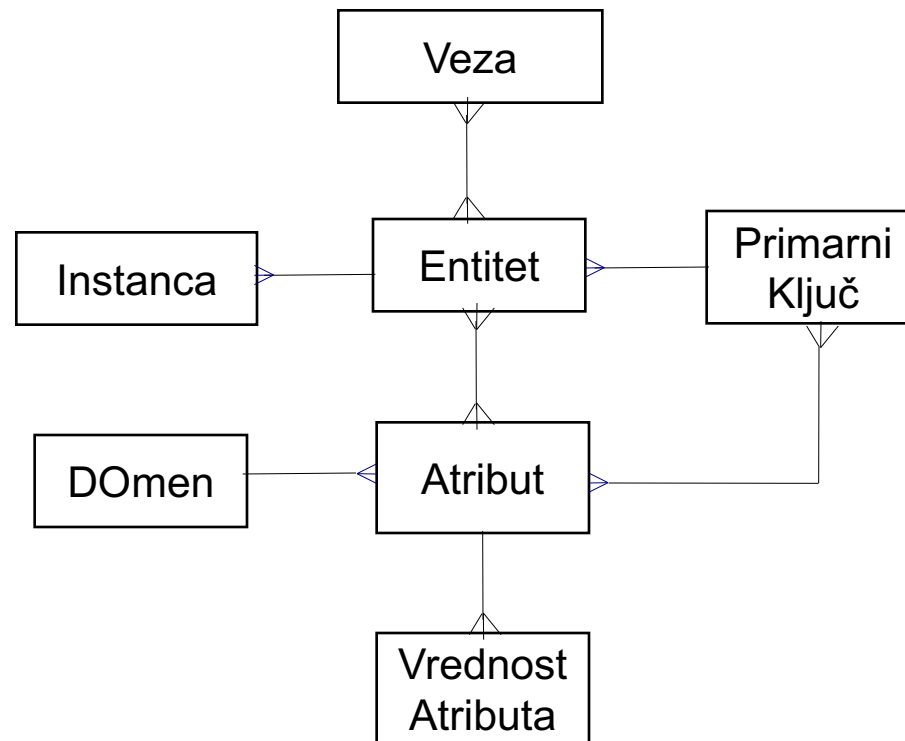
Otkrivanje Entiteta, Atributa, i Veza

- Pratiti imenice, imeničke fraze, glagole i prideve koje koriste stručnjaci u oblasti
- Imenice su potencijalni entiteti i atributi, ili instance i vrednosti atributa
- Pridevi su obično potencijalni atributi
- Glagoli naglašavaju potencijalne veze

Dodatne Smernice

- Smestite reči u kontekst da bi ih definisali
- Potražiti podatke iz nekoliko izvora
 - Ne oslanjati se samo na jednog stručnjaka
- Imajte šablone na umu

Model podataka od termina modelovanja



Tipovi Modela Podataka

- Konceptualni
 - Na visokom nivou
- Logički
 - Detaljniji
 - Nema veze više-prema-više
 - Spreman za fizičku implementaciju
- Fizički
 - Dodatni DBMS detalji
 - Spreman za implementiranje baze podataka

Normalizacija

Proces identifikacije jednog i najboljeg mesta za podatak

- Prva Normalna Forma

Red je u prvoj normalnoj formi samo ukoliko svi atributi sadrže samo atomske vrednosti

- Druga Normalna Forma

Red je u drugoj normalnoj formi akko je u prvoj normalnoj formi i ako svaki atribut koji nije deo ključa zavisi od celog ključa

- Treća Normalna Forma

Red je u trećoj normalnoj formi akko je u drugoj normalnoj formi i ako ni jedan ne ključni atribut ne zavisi od drugog ne ključnog atributa.

Ne-normalizovani podaci o STUDENT-u

StudentID	StudentName	MajorID	StudentMajor	CourseNum	CourseName	CourseCompDate
2907	Smith, Jacob R	MAT	Mathematics	MAT0011	Discrete Math	2002-08-01
				MAT0027	Calculus I	2002-04-30
				EGL0010	English Classics I	2001-12-30
4019	Patterson, Jane K	PHI	Philosophy	PHI0010	Intro to Philosophy	2002-04-30
				CS00100	Programming Languages	2002-04-30
5145	Neeld, Norris B	EGL	English Literature	SOC0102	Ascent of Man	2002-08-01
6132	Morrison, Xavier Q	MUS	Music	MUS0002	Origin of Jazz	2002-04-30
				SOC0102	Ascent of Man	2002-08-01
7810	Brown, Richard E	CS	Computer Science			
8966	Juarez, Samantha	EGL	English Literature	EGL0010	English Classics I	2001-12-30
				EGL0101	Shakespeare II	2002-08-01

Prva Normalna Forma

STUDENT

StudentID	LastName	FirstName	MiddleInit	MajorID	StudentMajor
2907	Smith	Jacob	R	MAT	Mathematics
4019	Patterson	Jane	K	PHI	Philosophy
5145	Neeld	Norris	B	EGL	English Literature
6132	Morrison	Xavier	Q	MUS	Music
7810	Brown	Richard	E	CS	Computer Science
8966	Juarez	Samantha		EGL	English Literature

KURS

StudentID	CourseNum	CourseName	CourseCompDate
2907	MAT0011	Discrete Math	2002-08-01
2907	MAT0027	Calculus I	2002-04-30
2907	EGL0010	English Classics I	2001-12-30
4019	PHI0010	Intro to Philosophy	2002-04-30
4019	CS00100	Programming Languages	2002-04-30
5145	SOC0102	Ascent of Man	2002-08-01
6132	MUS0002	Origin of Jazz	2002-04-30
6132	SOC0102	Ascent of Man	2002-08-01
8966	EGL0010	English Classics I	2001-12-30
8966	EGL0101	Shakespeare II	2002-08-01

Druga Normalna Forma

STUDENT

StudentID	LastName	FirstName	MiddleInit	MajorID	StudentMajor
2907	Smith	Jacob	R	MAT	Mathematics
4019	Patterson	Jane	K	PHI	Philosophy
5145	Neeld	Norris	B	EGL	English Literature
6132	Morrison	Xavier	Q	MUS	Music
7810	Brown	Richard	E	CS	Computer Science
8966	Juarez	Samantha		EGL	English Literature

ENROLLMENT

COURSE

CourseNum	CourseName	Credits
MAT0011	Discrete Math	3
MAT0027	Calculus I	4
EGL0010	English Classics I	3
PHI0010	Intro to Philosophy	3
CS00100	Programming Languages	3
SOC0102	Ascent of Man	3
MUS0002	Origin of Jazz	3

StudentID	CourseNum	CourseCompDate
2907	MAT0011	2002-08-01
2907	MAT0027	2002-04-30
2907	EGL0010	2001-12-30
4019	PHI0010	2002-04-30
4019	CS00100	2002-04-30
5145	SOC0102	2002-08-01
6132	MUS0002	2002-04-30
6132	SOC0102	2002-08-01
8966	EGL0010	2001-12-30
8966	EGL0101	2002-08-01

Treća Normalna Forma

STUDENT

StudentID	LastName	FirstName	MiddleInit	MajorID
2907	Smith	Jacob	R	MAT
4019	Patterson	Jane	K	PHI
5145	Neeld	Norris	B	EGL
6132	Morrison	Xavier	Q	MUS
7810	Brown	Richard	E	CS
8966	Juarez	Samantha		EGL

MAJOR

MajorID	StudentMajor
MAT	Mathematics
PHI	Philosophy
EGL	English Literature
MUS	Music
CS	Computer Science

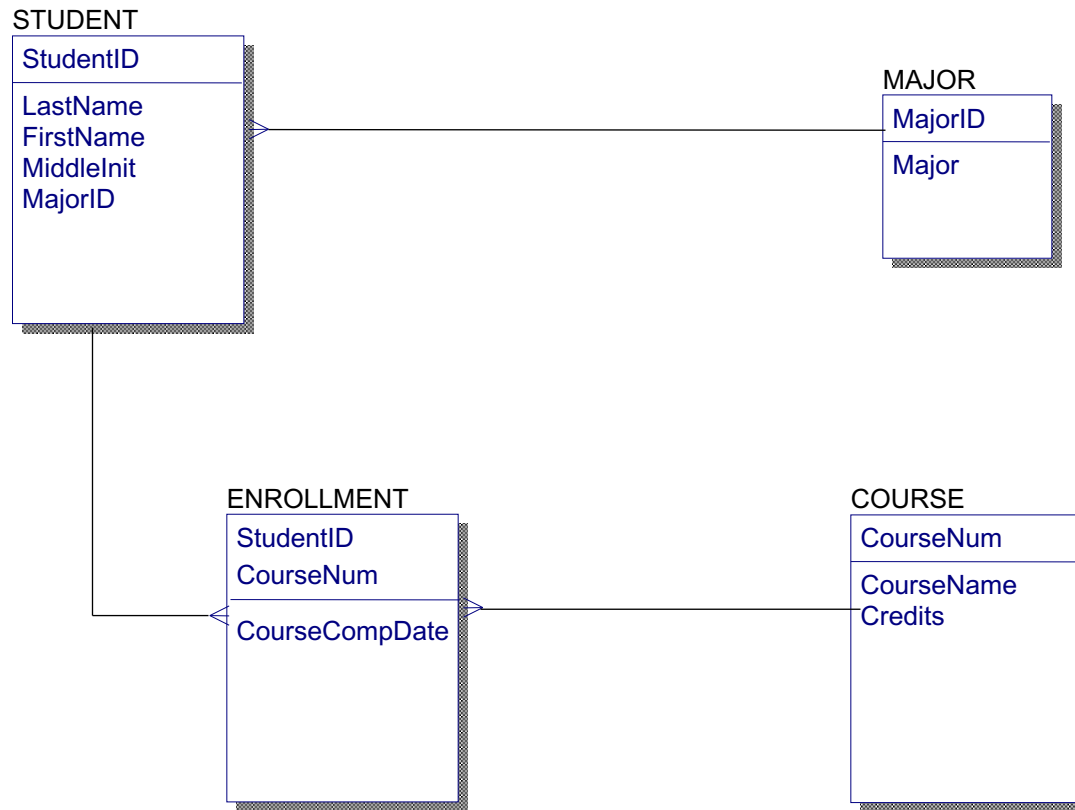
COURSE

CourseNum	CourseName	Credits
MAT0011	Discrete Math	3
MAT0027	Calculus I	4
EGL0010	English Classics I	3
PHI0010	Intro to Philosophy	3
CS00100	Programming Languages	3
SOC0102	Ascent of Man	3
MUS0002	Origin of Jazz	3

ENROLLMENT

StudentID	CourseNum	CourseCompDate
2907	MAT0011	2002-08-01
2907	MAT0027	2002-04-30
2907	EGL0010	2001-12-30
4019	PHI0010	2002-04-30
4019	CS00100	2002-04-30
5145	SOC0102	2002-08-01
6132	MUS0002	2002-04-30
6132	SOC0102	2002-08-01
8966	EGL0010	2001-12-30
8966	EGL0101	2002-08-01

Model podataka u Trećoj Normalnoj Formi



Još Normalnih Formi

- *Boyce Codd normalna form (BCNF)*
 - Dalje prečišćavanje 3NF. U njegovim kasnijim pisanjima Codd govori o BCNF kao 3NF. Red je u Boyce Codd normalnoj forma akko je svaka determinanta (atribut čija vrednost određuje druge vrednosti) u isto vreme i ključ kandidat. Najveći broj entiteta u 3NF su već u BCNF.
- *Četvrta normalna forma (4NF)*
 - Ni jedan etitet ne može da ima više od jedne jedan-prema-više veze ukoliko su atributi u tim vezama nezavisni jedan od drugog. Entitet je u 4NF akko je u 3NF i nema više setova multivrednosne zavisnosti.
- *Peta normalna form (5NF)*
 - Svaka join zavisnost entiteta mora da bude posledica njegovih kandidat ključeva.

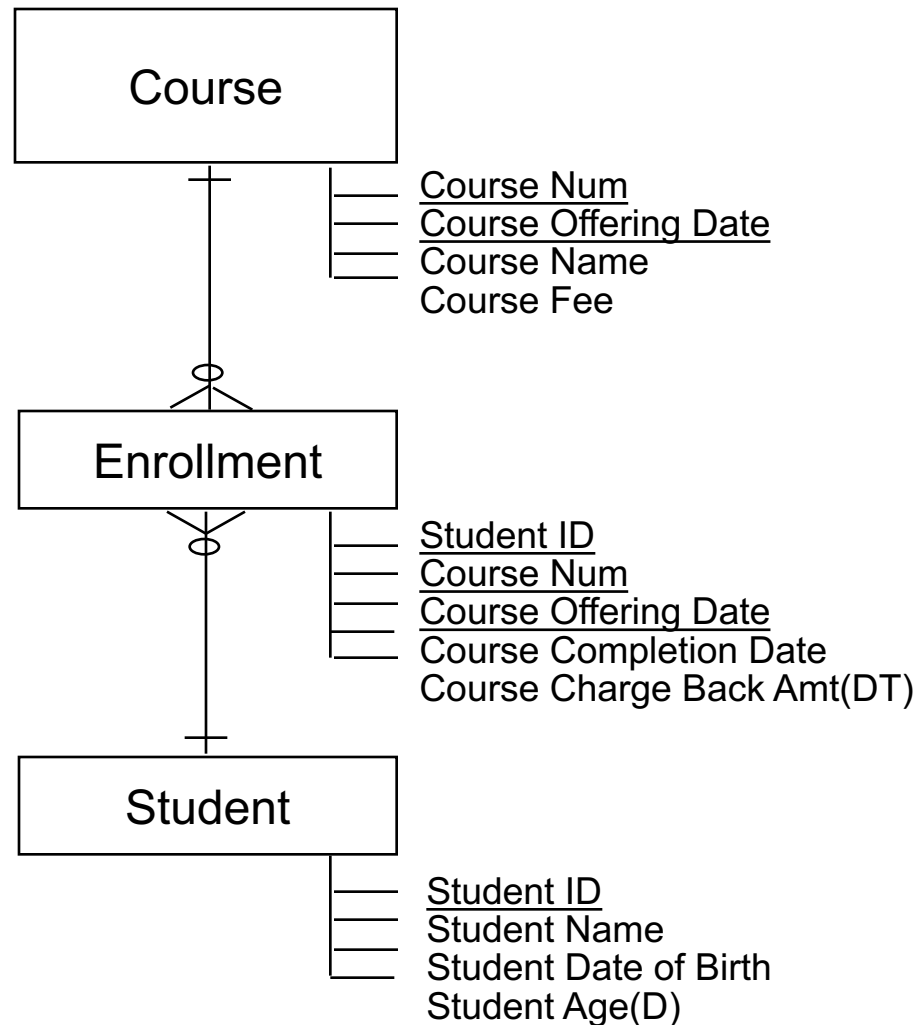
Normalizacija u Praksi

- Normalizacija osigurava da smo kreirali nacrt podataka za najbolji mogući dizajn podataka
- Normalizacija je logički proces.
 - Mnogo puta modelatori podataka pokušavaju da pretvore logički model u fizički, bez promena. Nije pametno forsirati DBA da se odreknu fizičkog modela
- U suštini, jedan na jedan, translacija logičkog u fizički model bi radila, ali to nije praktično, iz mnogih razloga.
- Jedina normalna forma koja je obavezna je 1NF. 1NF obezbeđuje tabularne podatke.

Dodatni Problemi u Modelovanju Podataka

- Podtipovi i super-tipovi
- Upravljanje vremenom
- Jedan-prema-jedan veze?
- Izvedeni podaci (videti sledeći slajd)

Izvedeni i osnovni podaci



Pitanja

