



Bazat e të dhënave

Pjesa 6 – Normalizimi

Prof. Asoc. Dr. Ermir Rogova



Objektivat

- Pas përfundimit të këtij kapitulli, do të jeni në gjendje të:
 - Shpjegoni normalizimin dhe rolin e saj në procesin e dizajnit të bazës së të dhënave
 - Identifikoni dhe përshkruani secilën nga format normale: 1NF, 2NF, 3NF, BCNF dhe 4NF
 - Shpjegoni se si format normale mund të shndërrohen nga format më të ulta normale në format më të larta normale
 - Zbatoni rregullat e normalizimit për të vlerësuar dhe korrigjuar strukturat e tabelës
 - Identifikoni situatat që kërkojnë denormalizim për të gjeneruar informacion në mënyrë efikase
 - Përdorni një listë kontrolli për modelimin e të dhënave për të kontrolluar që ERD plotëson një sërë kërkesash minimale



Tabelat e të dhënave dhe normalizimi (1/2)

- Normalizimi: vlerësimi dhe korigjimi i strukturave të tabelës për të minimizuar tepricat (redudancën) e të dhënave
 - Redukton anomalitë e të dhënave
 - Cakton attribute në tabela bazuar në përcaktimin
- Format normale
 - Forma e parë normale (1NF)
 - Forma e dytë normale (2NF)
 - Forma e tretë normale (3NF)



Tabelat e të dhënave dhe normalizimi (2/2)

- Pikëpamja strukturore e formave normale
 - Format më të larta normale janë më të mira se format më të ulëta
 - Strukturat e dizajnuara në mënyrë korrekte 3NF plotësojnë kërkesat e formës së katërt normale (4NF)
- Denormalizimi: prodhon një formë më të ulët normale
 - Rezultatet në rritjen e performancës dhe tepricë më të madhe të të dhënave



Nevoja për normalizim

- Përdoret gjatë dizajnit të një strukture të re të bazës së të dhënave
 - Analizon marrëdhëniet midis attributeve brenda secilit entitet
 - Përcakton nëse struktura mund të përmirësohet përmes normalizimit
 - Përmirëson strukturën ekzistuese të të dhënave dhe krijon një dizajn të duhur të bazës së të dhënave



Procesi i normalizimit (1/5)

- Objektivi është të sigurojmë që secila tabelë përputhet me konceptin e marrëdhënieve të formuara mirë
 - Secila tabelë përfaqëson një entitet të vetëm
 - Çdo kryqëzim i rreshtave / kolonave përmban vetëm një vlerë dhe jo një grup vlerash
 - Asnjë artikull i të dhënave nuk do të ruhet në mënyrë të panevojshme në më shumë se një tabelë
 - Të gjitha atributet jo prime në një tabelë varen nga çelësi kryesor
 - Çdo tabelë nuk ka anomali të futjes, azhurnimit ose fshirjes



Procesi i normalizimit (2/5)

- Siguron që të gjitha tabelat të jenë të paktën në 3NF
 - Forma më të larta nuk ka të ngjarë të hasen në mjedisin e biznesit
- Përpunohen tabelat një-nga-një
 - Identifikohen varësitë
 - Në mënyrë progresive shpërbëhet tabela në grupe tabelash me marrëdhënie ndërmejt tyre



Procesi i normalizimit (3/5)

Tabla 6.2: Fromat Normale		
Forma normale	Karakteristika	Seksioni
Forma e pare normale First normal form (1NF)	Fromat tabelor, pa grupe përsëritëse, dhe çelësat primarë të identifikuar	6-3a
Forma e dyte normale Second normal form (2NF)	1NF dhe nuk përmban mvarshmëri parciale	6-3b
Forma tretë normale Third normal form (3NF)	2NF dhe nuk përmban mvarshmëri tranzitive	6-3c
Boyce-Codd normal form (BCNF)	Çdo determinantë është një çelës kandidat (rasti special i 3NF)	6-6a
Forma e katërt normale Fourth normal form (4NF)	3NF dhe nuk përmban mvarshëmëri shumëvlerëshe të pavarura	6-6b



Procesi i normalizimit (4/5)

**Tabela 6.3: Konceptet e Varësisë
Funksionale**

Koncepti	Përkufizimi
Varësia funksionale	Atributi B është plotësisht i varur funksionalisht nga atributi A nëse secila vlerë e A përcakton një dhe vetëm një vlerë të B.
Varësia funksionale (përkufizimi i përgjithësuar)	Atributi A përcakton atributin B (domethënë, B është funksionalisht i varur nga A) nëse të gjithë rreshtat në tabelë që bien dakord në vlerë për atributin A pajtohen gjithashtu në vlerë për atributin B.
Varësia plotësisht funksionale (çelësi i përbërë)	Nëse atributi B është i varur funksionalisht nga një çelës i përbërë A por jo nga ndonjë nëngrup i këtij çelësi të përbërë, atributi B është plotësisht i varur funksionalisht nga A.

Procesi i normalizimit (5/5)

- Varësia e pjesëshme: varësia funksionale në të cilën determinanta është vetëm pjesë e çelësit primar
 - Supozimi: një çelës kandidat
 - Lehtë për tu identifikuar
- Varësia tranzitive: atributi varet nga një atribut tjetër që nuk është pjesë e çelësit primar
 - Më e vështirë për t'u identifikuar midis një grupi të dhënash
 - Ndodh vetëm kur ekziston një varësi funksionale midis attributeve jo primare



Konvertimi në Formën e Parë Normale (1NF) (1/3)

- Procedura me tre hapa
 - Eliminoni atributet me shumë vlera
 - Identifikoni çelësin kryesor
 - Identifikoni të gjitha varësitë
- Diagrami i varësisë: përshkruan të gjitha varësitë e gjetura brenda strukturës së dhënë të tabelës
 - Ndhmon për të marrë një pasqyrë të të gjitha marrëdhënieve midis attributeve të tabelës
 - E bën më pak të mundshme që një varësi e rëndësishme të neglizhohet

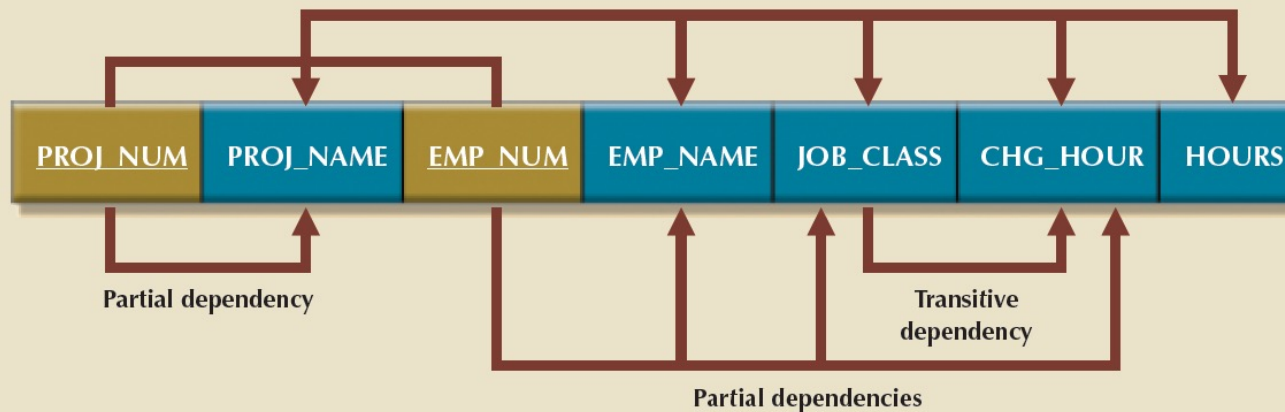


Konvertimi në Formën e Parë Normale (1NF) (2/3)

- 1NF përshkruan formatin tabelor në të cilin:
 - Të gjitha atributet kryesore janë të përcaktuara
 - Nuk ka grupe përsëritëse në tabelë
 - Të gjitha atributet janë të varura nga çelësi kryesor
- Të gjitha tabelat relacionale plotësojnë kërkesat e 1NF
- Disa tabela përmbajnë varësi të pjeshme

Konvertimi në Formën e Parë Normale (1NF) (3/3)

FIGURE 6.3 FIRST NORMAL FORM (1NF) DEPENDENCY DIAGRAM



1NF (PROJ_NUM, EMP_NUM, PROJ_NAME, EMP_NAME, JOB_CLASS, CHG_HOURS, HOURS)

PARTIAL DEPENDENCIES:

(PROJ_NUM \Rightarrow PROJ_NAME)

(EMP_NUM \Rightarrow EMP_NAME, JOB_CLASS, CHG_HOUR)

TRANSITIVE DEPENDENCY:

(JOB_CLASS \Rightarrow CHG_HOUR)

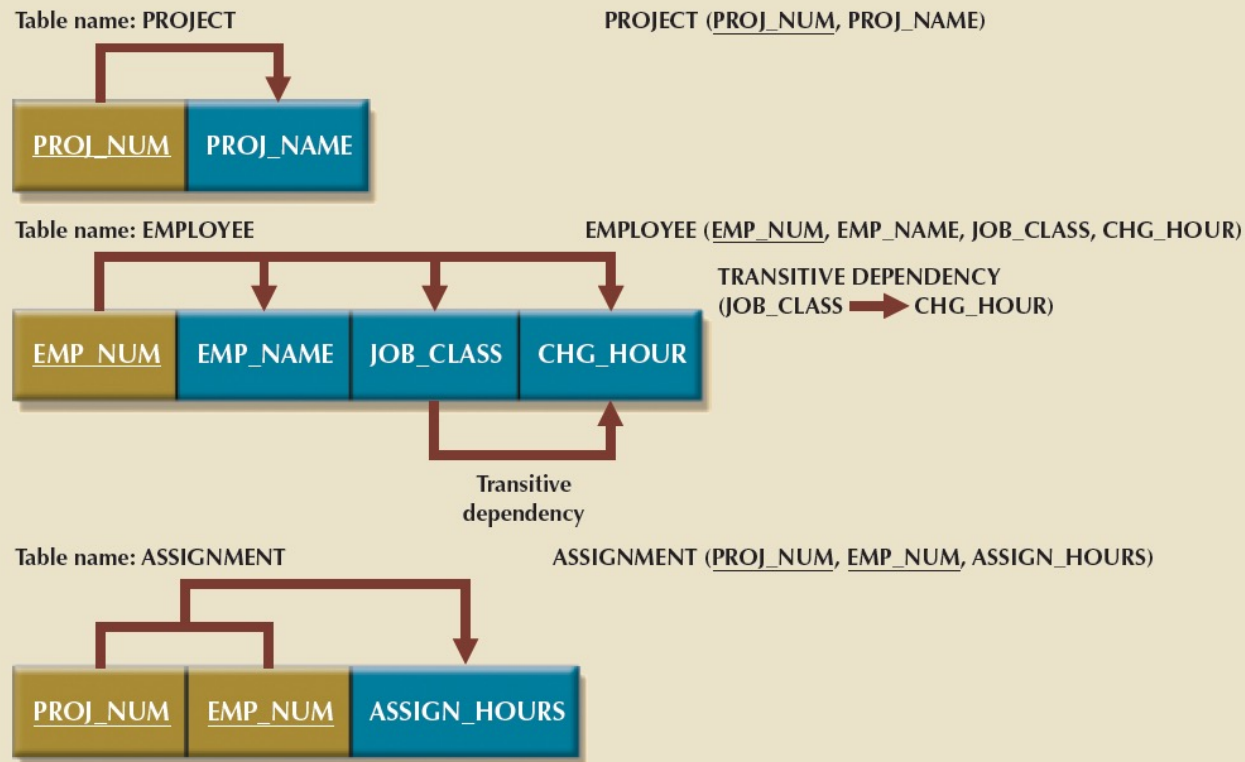


Konvertimi në Formën e Dytë Normale (2NF) (1/2)

- Konvertimi në 2NF ndodh vetëm kur 1NF ka një çelës primar të përbërë
 - Nëse 1NF ka një çelës primar me një atribut të vetëm, atëherë tabela është automatikisht në 2NF
- Shndërrimi 1NF-në-2NF është i thjeshtë
 - Bëni tabela të reja për të eliminuar varësitë e pjeshme
 - Ricaktoni atributet përkatëse të varura
- Tabela është në 2NF kur:
 - Është në 1NF
 - Nuk përfshin asnjë varësi të pjeshme

Konvertimi në Formën e Dytë Normale (2NF) (2/2)

FIGURE 6.4 SECOND NORMAL FORM (2NF) CONVERSION RESULTS



Konvertimi në Formën e Tretë Normale (3NF) (1/2)

- Anomalitë e të dhënave të krijuara nga organizata e bazës së të dhënave të paraqitura në Figurën 6.4 eliminohen lehtësisht
 - Bëni tabela të reja për të eleminuar varësitë tranzitore
 - Ricaktoni atributet përkatëse të varura
- Tabela është në 3NF kur:
 - Është në 2NF
 - Nuk përmban asnjë varësi kalimtare

Konvertimi në Formën e Tretë Normale (3NF) (2/2)

FIGURE 6.5 THIRD NORMAL FORM (3NF) CONVERSION RESULTS



Table name: PROJECT

PROJECT (PROJ_NUM, PROJ_NAME)

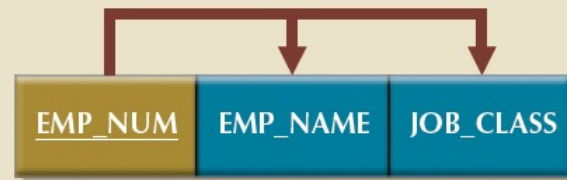


Table name: EMPLOYEE

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)



Table name: JOB

JOB (JOB_CLASS, CHG_HOUR)



Table name: ASSIGNMENT

ASSIGNMENT (PROJ_NUM, EMP_NUM, ASSIGN_HOURS)

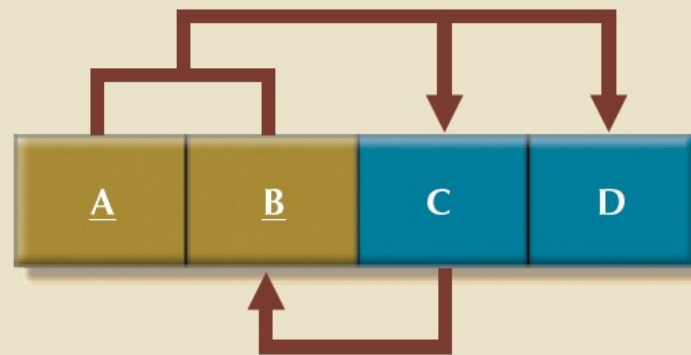


Forma normale Boyce-Codd (1/4)

- Çdo determinantë në tabelë duhet të jetë një çelës kandidat
 - Çelësi kandidat: të njëjtat karakteristika si çelësi primar por nuk janë zgjedhur të jenë çelësi primar
 - Ekuivalente me 3NF kur tabela përmban vetëm një çelës kandidat
 - Shkelet vetëm kur tabela përmban më shumë se një çelës kandidat
 - Konsiderohet të jetë një rast i veçantë i 3NF

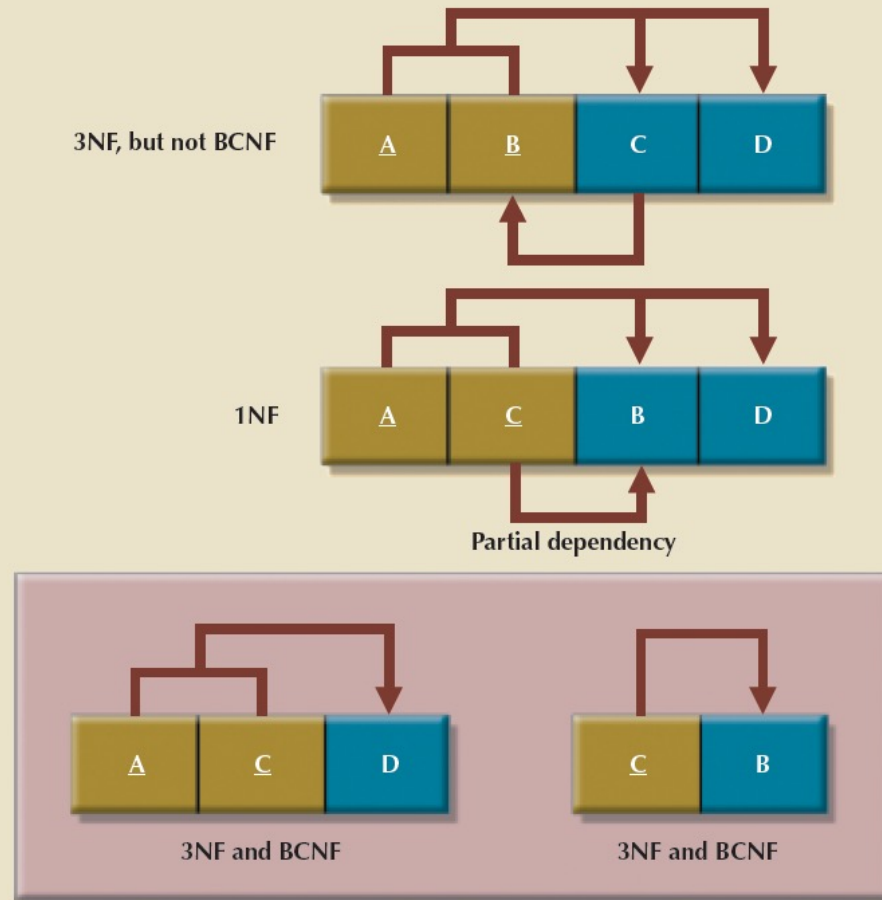
Forma normale Boyce-Codd (1/4)

FIGURE 6.8 A TABLE THAT IS IN 3NF BUT NOT IN BCNF



Forma normale Boyce-Codd (2/4)

FIGURE 6.9 DECOMPOSITION TO BCNF





Forma e Katërt Normale (4NF) (1/2)

- Rregullat
 - Të gjitha atributet duhet të jenë të varura nga çelësi primar, por ato duhet të jenë të pavarura nga njëra-tjetra
 - Asnjë rresht nuk mund të përmbajë dy ose më shumë fakte shumëvlerëshe rreth një entiteti
- Tabela është në 4NF kur:
 - Është në 3NF
 - Nuk ka varësi shumëvlerëshe



Forma e Katërt Normale (4NF) (2/2)

FIGURE 6.11 TABLES WITH MULTIVALUED DEPENDENCIES

Database name: Ch06_Service

Table name: VOLUNTEER_V1

EMP_NUM	ORG_CODE	ASSIGN_NUM
10123	RC	1
10123	UW	3
10123		4

Table name: VOLUNTEER_V2

EMP_NUM	ORG_CODE	ASSIGN_NUM
10123	RC	
10123	UW	
10123		1
10123		3
10123		4

Table name: VOLUNTEER_V3

EMP_NUM	ORG_CODE	ASSIGN_NUM
10123	RC	1
10123	RC	3
10123	UW	4



Normalizimi dhe dizajni i databazave (1/6)

- Normalizimi duhet të jetë pjesë e procesit të dizajnit
 - Entitetet e propozuara duhet të plotësojnë formën normale të kërkuar para se të krijohen strukturat tabelore
- Principet dhe procedurat e normalizimit që duhen kuptuar për të ridizajnuar dhe modifikuar bazat e të dhënave
 - ERD krijohet përmes një procesi përsëritës
 - Normalizimi përqendrohet në karakteristikat e entiteteve specifike

Normalizimi dhe dizajni i databazave (2/6)

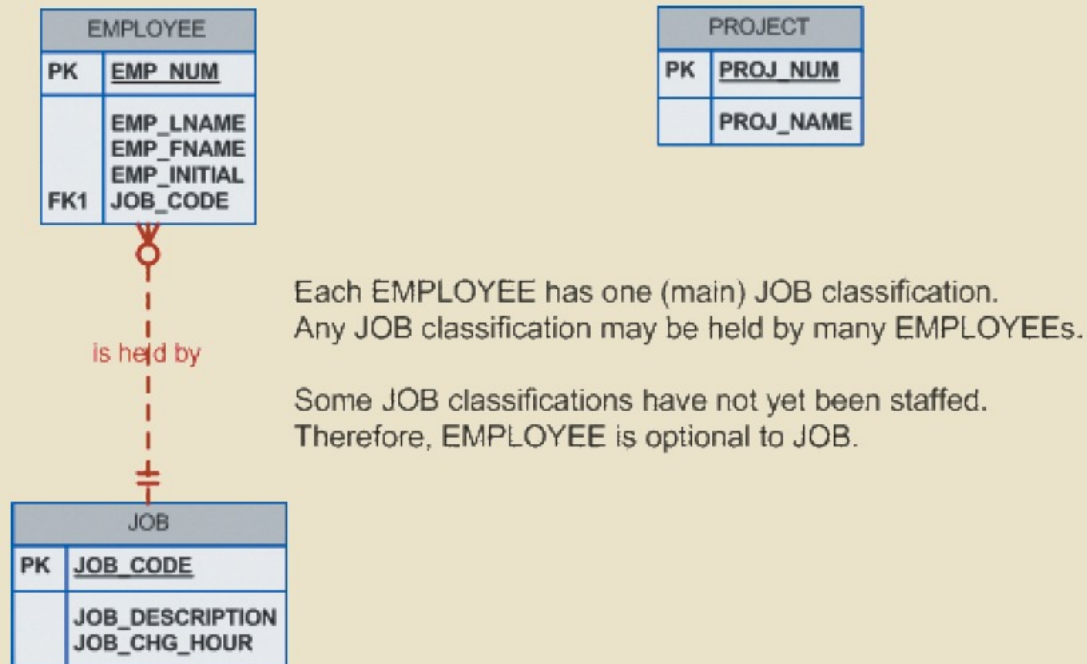
FIGURE 6.13 INITIAL CONTRACTING COMPANY ERD

EMPLOYEE	
PK	<u>EMP_NUM</u>
	EMP_LNAME
	EMP_FNAME
	EMP_INITIAL
	JOB_DESCRIPTION
	JOB_CHG_HOUR

PROJECT	
PK	<u>PROJ_NUM</u>
	PROJ_NAME

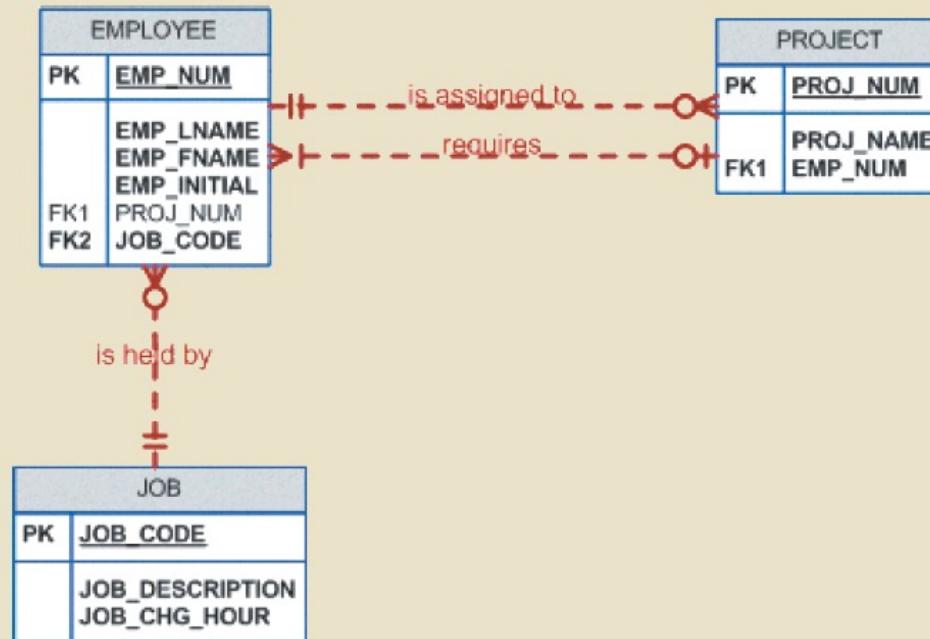
Normalizimi dhe dizajni i databazave (3/6)

FIGURE 6.14 MODIFIED CONTRACTING COMPANY ERD



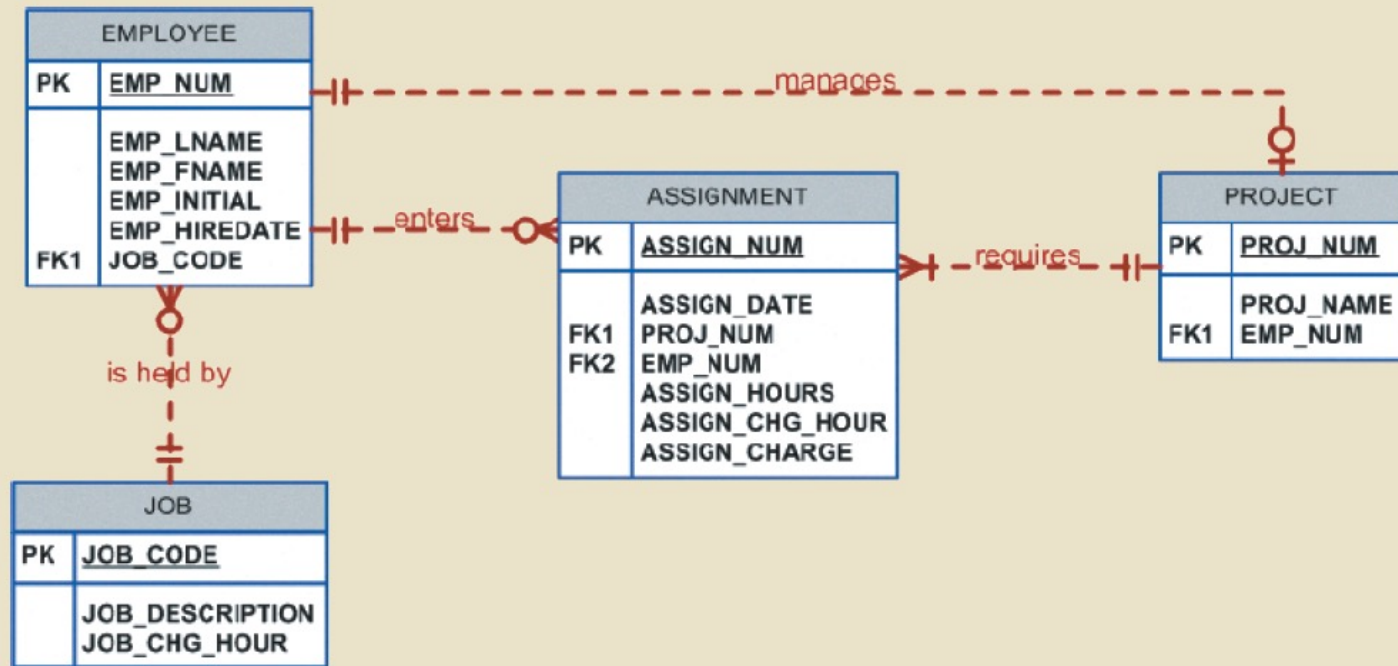
Normalizimi dhe dizajni i databazave (4/6)

FIGURE 6.15 INCORRECT M:N RELATIONSHIP REPRESENTATION



Normalizimi dhe dizajni i databazave (5/6)

FIGURE 6.16 FINAL CONTRACTING COMPANY ERD



Normalizimi dhe dizajni i databazave (6/6)

FIGURE 6.17 THE IMPLEMENTED DATABASE

Table name: EMPLOYEE

EMP_NUM	EMP_LNAME	EMP_FNAME	EMP_INITIAL	EMP_HIREDATE	JOB_CODE
101	News	John	G	08-Nov-00	502
102	Senior	David	H	12-Jul-89	501
103	Arbough	June	E	01-Dec-97	503
104	Ramoras	Anne	K	15-Nov-88	501
105	Johnson	Alice	K	01-Feb-94	502
106	Smithfield	William		22-Jun-05	500
107	Alonzo	Maria	D	10-Oct-94	500
108	Washington	Ralph	B	22-Aug-89	501
109	Smith	Larry	W	18-Jul-99	501
110	Olenko	Gerald	A	11-Dec-96	505
111	Wabash	Geoff	B	04-Apr-89	506
112	Smithson	Darlene	M	23-Oct-95	507
113	Joenbrood	Delbert	K	15-Nov-94	508
114	Jones	Annelise		20-Aug-91	508
115	Bawangi	Travis	B	25-Jan-90	501
116	Pratt	Gerald	L	05-Mar-95	510
117	Williamson	Angie	H	19-Jun-94	509
118	Frommer	James	J	04-Jan-06	510

Database name: Ch06_ConstructCo

Table name: JOB

JOB_CODE	JOB_DESCRIPTION	JOB_CHG_HOUR
500	Programmer	35.75
501	Systems Analyst	96.75
502	Database Designer	105.00
503	Electrical Engineer	84.50
504	Mechanical Engineer	67.90
505	Civil Engineer	55.78
506	Clerical Support	26.87
507	DSS Analyst	45.95
508	Applications Designer	48.10
509	Bio Technician	34.55
510	General Support	18.36

Table name: PROJECT

PROJ_NUM	PROJ_NAME	EMP_NUM
15	Evergreen	105
18	Amber Wave	104
22	Rolling Tide	113
25	Starflight	101

Table name: ASSIGNMENT

ASSIGN_NUM	ASSIGN_DATE	PROJ_NUM	EMP_NUM	ASSIGN_HOURS	ASSIGN_CHG_HOUR	ASSIGN_CHARGE
1001	04-Mar-16	15	103	2.6	84.50	219.70
1002	04-Mar-16	18	118	1.4	18.36	25.70
1003	05-Mar-16	15	101	3.6	105.00	378.00
1004	05-Mar-16	22	113	2.5	48.10	120.25
1005	05-Mar-16	15	103	1.9	84.50	160.55
1006	05-Mar-16	25	115	4.2	96.75	406.35
1007	05-Mar-16	22	105	5.2	105.00	546.00
1008	05-Mar-16	25	101	1.7	105.00	178.50
1009	05-Mar-16	15	105	2.0	105.00	210.00
1010	06-Mar-16	15	102	3.8	96.75	367.65
1011	06-Mar-16	22	104	2.6	96.75	251.55
1012	06-Mar-16	15	101	2.3	105.00	241.50
1013	06-Mar-16	25	114	1.8	48.10	86.58
1014	06-Mar-16	22	111	4.0	26.87	107.48
1015	06-Mar-16	25	114	3.4	48.10	163.54
1016	06-Mar-16	18	112	1.2	45.95	55.14
1017	06-Mar-16	18	118	2.0	18.36	36.72
1018	06-Mar-16	18	104	2.6	96.75	251.55
1019	06-Mar-16	15	103	3.0	84.50	253.50
1020	07-Mar-16	22	105	2.7	105.00	283.50
1021	08-Mar-16	25	108	4.2	96.75	406.35
1022	07-Mar-16	25	114	5.8	48.10	278.98
1023	07-Mar-16	22	106	2.4	35.75	85.80

Denormalizimi (1/2)

- Qëllimet e dizajnit
 - Krijimi i marrëdhënieve të normalizuara
 - Përpunimi i kërkesave dhe shpejtësia
- Numri i tabelave të bazës së të dhënave rritet
 - Tabelat janë zbërthyer në përputhje me kërkesat e normalizimit
- Bashkimi i një numri më të madh të tabelave
 - Merr operacione shtesë të hyrjes/daljes (I/O) dhe logjikës së procesimit
 - Ul shpejtësinë e sistemit
- Defektet në tabelat e panormalizuara
 - Përditësimet e të dhënave janë më pak efikase sepse tabelat janë më të mëdha
 - Indeksimi është më i rëndë
 - Nuk ka strategji të thjeshtë për krijimin e tabelave virtuale të njohura si pamje (view)





Denormalizimi (2/2)

Table 6.6: Common Denormalization Examples		
Case	Example	Rationale and Controls
Redundant data	Storing ZIP and CITY attributes in the AGENT table when ZIP determines CITY (see Figure 2.2)	Avoid extra join operations Program can validate city (drop-down box) based on the zip code
Derived data	Storing STU_HRS and STU_CLASS (student classification) when STU_HRS determines STU_CLASS (see Figure 3.28)	Avoid extra join operations Program can validate classification (lookup) based on the student hours
Preaggregated data (also derived data)	Storing the student grade point average (STU_GPA) aggregate value in the STUDENT table when this can be calculated from the ENROLL and COURSE tables (see Figure 3.28)	Avoid extra join operations Program computes the GPA every time a grade is entered or updated STU_GPA can be updated only via administrative routine
Information requirements	Using a temporary denormalized table to hold report data; this is required when creating a tabular report in which the columns represent data that are stored in the table as rows (see Figures 6.17 and 6.18)	Impossible to generate the data required by the report using plain SQL No need to maintain table Temporary table is deleted once report is done Processing speed is not an issue

Lista kontrolluese e modelimit të të dhënave (1/6)

- Rregullat e biznesit
 - Dokumentoni dhe verifikoni siç duhet të gjitha rregullat e biznesit me përdoruesit fundorë
 - Sigurohuni që të gjitha rregullat e biznesit janë shkruar saktësisht, qartë dhe thjesht
 - Rregullat e biznesit duhet të ndihmojnë në identifikimin e entiteteve, attributeve, marrëdhënieve dhe kufizimeve
 - Identifikoni burimin e të gjitha rregullave të biznesit dhe siguroni që secila rregull biznesi është e justifikuar, e datuar dhe e nënshkruar nga një autoritet miratues



Lista kontrolluese e modelimit të të dhënave (2/6)

- Modelimi i të dhënave
 - Konventat e emërimit: të gjithë emrat duhet të jenë të kufizuar në gjatësi (madhësia varet nga baza e të dhënave)
- Emrat e entiteteve:
 - Duhet të jenë emra që janë të njohur për biznesin dhe duhet të jenë të shkurtër dhe kuptimplotë
 - Duhet të dokumentojë shkurtesat, sinonimet dhe pseudonimet për secilin entitet
 - Duhet të jetë unike brenda modelit
 - Për entitetet e përbëra, mund të përfshijë një kombinim të emrave të shkurtuar të entiteteve të lidhura përmes entitetit të përbërë



Lista kontrolluese e modelimit të të dhënave (3/6)

- Emrat e attributeve:
 - Duhet të jetë unike brenda entitetit
 - Duhet të përdorë shkurtesën e entitetit si parashtesë
 - Duhet të jetë përshkruese e karakteristikës
 - Duhet të përdorë prapashtesa të tilla si `_ID`, `_NUM`, ose `_CODE` për atributin PK
 - Nuk duhet të jetë një fjalë e rezervuar
 - Nuk duhet të përmbajë hapësira ose karaktere speciale si `@`, `!`, `Ose` &
- Emrat e marrëdhënieve:
 - Duhet të jenë folje aktive ose pasive që tregojnë qartë natyrën e marrëdhënies

Lista kontrolluese e modelimit të të dhënave (4/6)

- Entitetet:
 - Çdo entitet duhet të përfaqësojë një subjekt të vetëm
 - Çdo entitet duhet të përfaqësojë një set instancash të dallueshme të entitetit
 - Të gjitha entitetet duhet të jenë në 3NF ose më të larta
 - Çdo entitet nën 3NF duhet të justifikohet
 - Granuliteti i instancës së entitetit duhet të përcaktohet qartë
 - PK duhet të përcaktohet qartë dhe të mbështesë granulitetin e zgjedhur të të dhënave

Lista kontrolluese e modelimit të të dhënave (5/6)

- Atributet:
 - Duhet të jetë e thjeshta dhe me vlerë të vetme (të dhëna atomike)
 - Duhet të dokumentohen vlerat e paracaktuara, kufizimet, sinonimet dhe pseudonimet
 - Atributet e derivuara duhet të identifikohen qartë dhe të përfshijnë burimin (burimet)
 - Nuk duhet të jetë redundante nëse kjo nuk kërkohet për saktësinë e transaksionit, performancën ose mbajtjen e një historie
 - Atributet jo çelës duhet të jenë plotësisht të varura nga atributi PK
- Marrëdhëniet:
 - Duhet të identifikojnë qartë pjesëmarrësit e marrëdhënieve
 - Duhet të përcaktojnë qartë pjesëmarrjen, lidhshmërinë dhe kardinalitetin e dokumenteve





Lista kontrolluese e modelimit të të dhënave (6/6)

- Modeli ER:
 - Duhet të validohet kundrejt proceseve të pritshme: insertimet, azhurnimet dhe fshirjet
 - Duhet të vlerësojmë se ku, kur dhe si të mbajmë një histori
 - Nuk duhet të përmbajë marrëdhënie redundante përveç nëse kërkohet
 - Duhet të minimizojë redundancën e të dhënave për të siguruar azhurnime në një vend të vetëm
 - Duhet të jetë në përputhje me rregullin e të dhënave minimale: Gjithçka që nevojitet është aty, dhe gjithçka që është aty është e nevojshme



Përmbledhje (1/2)

- Normalizimi është një teknikë e përdorur për të dizajnuar tabela në të cilat minimizohen tepricat e të dhënave
- Një tabelë është në 1NF kur përcaktohen të gjitha atributet kryesore dhe të gjitha atributet e mbetura varen nga çelësi primar
- Një tabelë është në 2NF kur është në 1NF dhe nuk përmban varësi të pjesshme
- Një tabelë është në 3NF kur është në 2NF dhe nuk përmban varësi kalimtare
- Një tabelë që nuk është në 3NF mund të ndahet në tabela të reja derisa të gjitha tabelat të përmbushin kërkesat e 3NF



Përmbledhje (2/2)

- Normalizimi është një pjesë e rëndësishme - por vetëm një pjesë - e procesit të dizajnit
- Një tabelë në 3NF mund të përmbajë varësi shumëvlerëshe që prodhojnë ose vlera të shumta null ose të dhëna të tepërta
- Sa më i madh numri i tabelave, aq më shumë operacione shtesë I/O dhe logjikë të procesimit ju duhet për to bashkuar ato
- Lista kontrolluese e modelimit të të dhënave siguron një mënyrë që projektuesi të kontrollojë nëse ERD plotëson një set kërkesash minimale



Pyetje???