

به نام آنکه جان را فکرت آموخت



بخش دهم: نرمال تر سازی پایگاه داده

مرتضی امینی

نیمسال اول ۱۴۰۱-۱۴۰۰

(محتویات اسلایدها برگرفته از یادداشت‌های کلاسی استاد محمدتقی روحانی رانکوهی است.)



□ در طراحی پایگاه داده‌های رابطه‌ای باید موارد زیر را مشخص نمود:

□ مجموعه‌ای از رابطه‌ها

□ کلید(های) کاندید هر رابطه

□ کلید اصلی هر رابطه

□ کلیدهای خارجی هر رابطه (در صورت وجود)

□ محدودیت‌های جامعیتی ناظر بر هر رابطه

طراحی با روش بالا به پایین (Top-Down)

طراحی با روش سنتز [نرمال‌ترسازی رابطه‌ها]

□ روشهای طراحی RDB:



روش طراحی بالا به پایین

ابتدا مدل‌سازی داده‌ها را (با روش [E]ER یا UML) انجام می‌دهیم و سپس مدل‌سازی را به مجموعه‌ای از رابطه‌ها تبدیل می‌کنیم.

روش طراحی سنتز رابطه‌ای (نرمال‌سازی)

ابتدا مجموعه صفات خرد جهان واقع را مشخص می‌کنیم. سپس با تحلیل قواعد و محدودیت‌های ناظر به صفات و تشخیص وابستگی‌های بین آنها، صفات را متناسباً با هم سنتز می‌کنیم (نوعی گروه‌بندی) تا به مجموعه‌ای از رابطه‌های نرمال دست یابیم.

در عمل روش ترکیبی استفاده می‌شود، یعنی ابتدا روش بالا به پایین، سپس نرمال‌سازی.



□ روش طراحی بالا به پایین: تبدیل نمودار [E]ER به مجموعه‌ای از رابطه‌های نرمال (و نه لزوماً در

نرمال‌ترین صورت). در طراحی RDB، نهایتاً طراح تصمیم می‌گیرد چند رابطه داشته باشد و عنوان

(Heading) هر رابطه چه باشد.

□ در نمودار مدل‌سازی معنایی داده‌ها، حالات متعدد داریم، که به نحوه طراحی بر اساس آن در بخش‌های قبلی

اشاره شد.



طراحی RDB – روش سنتز یا نرمال‌ترسازی رابطه‌ها

۵

بخش دهم: نرمال‌ترسازی پایگاه داده

□ ایده اصلی در نرمال‌ترسازی: یک رابطه، هر چند نرمال (با تعریفی که قبلاً دیدیم) ممکن است آنومالی

(مشکل) داشته باشد در عملیات ذخیره‌سازی (در درج، حذف یا بهنگام‌سازی).

□ آنومالی در درج: عدم امکان درج یک فقره اطلاع که منطقاً باید قابل درج باشد.

□ آنومالی در حذف: حذف یک اطلاع ناخواسته در پی حذف اطلاع خواسته.

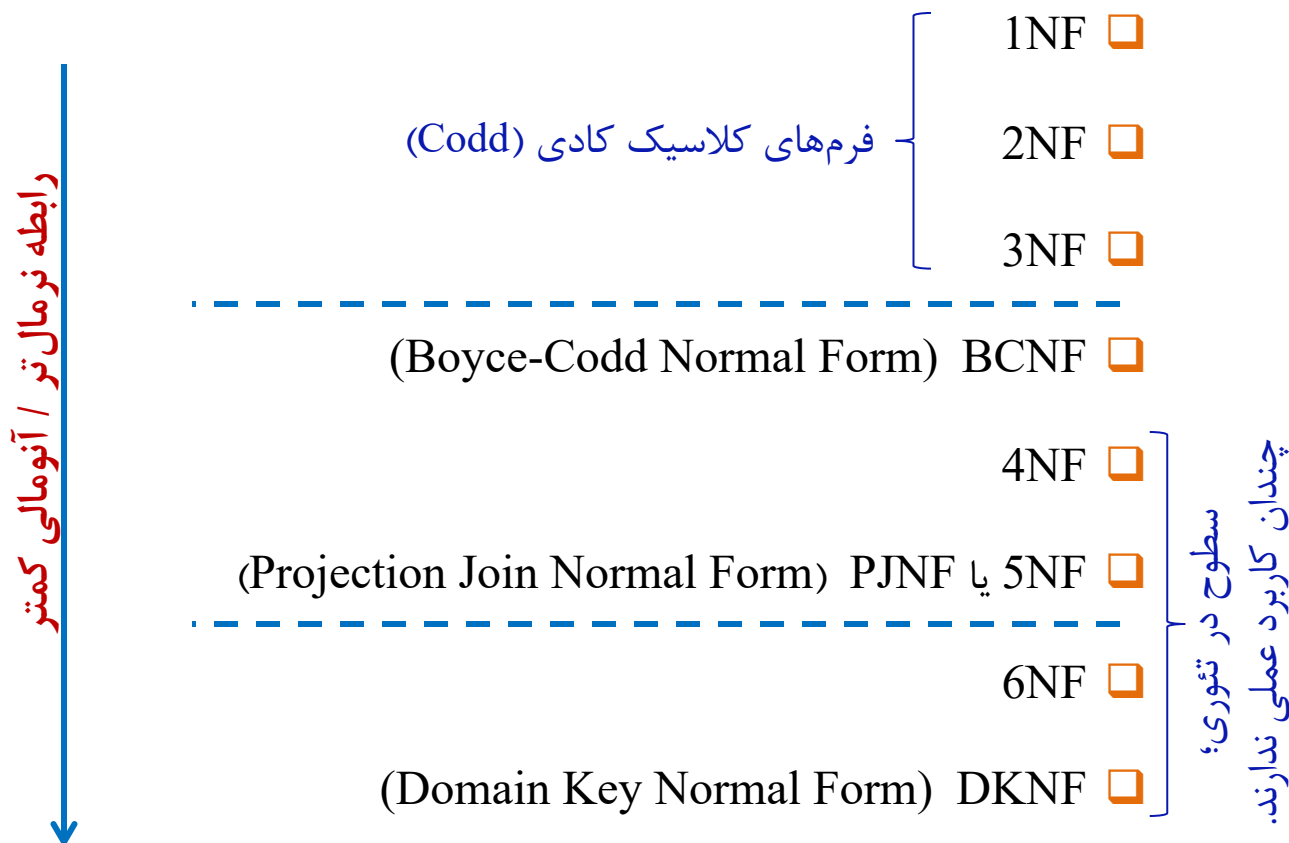
□ آنومالی در بهنگام‌سازی: بروز فزون‌کاری.

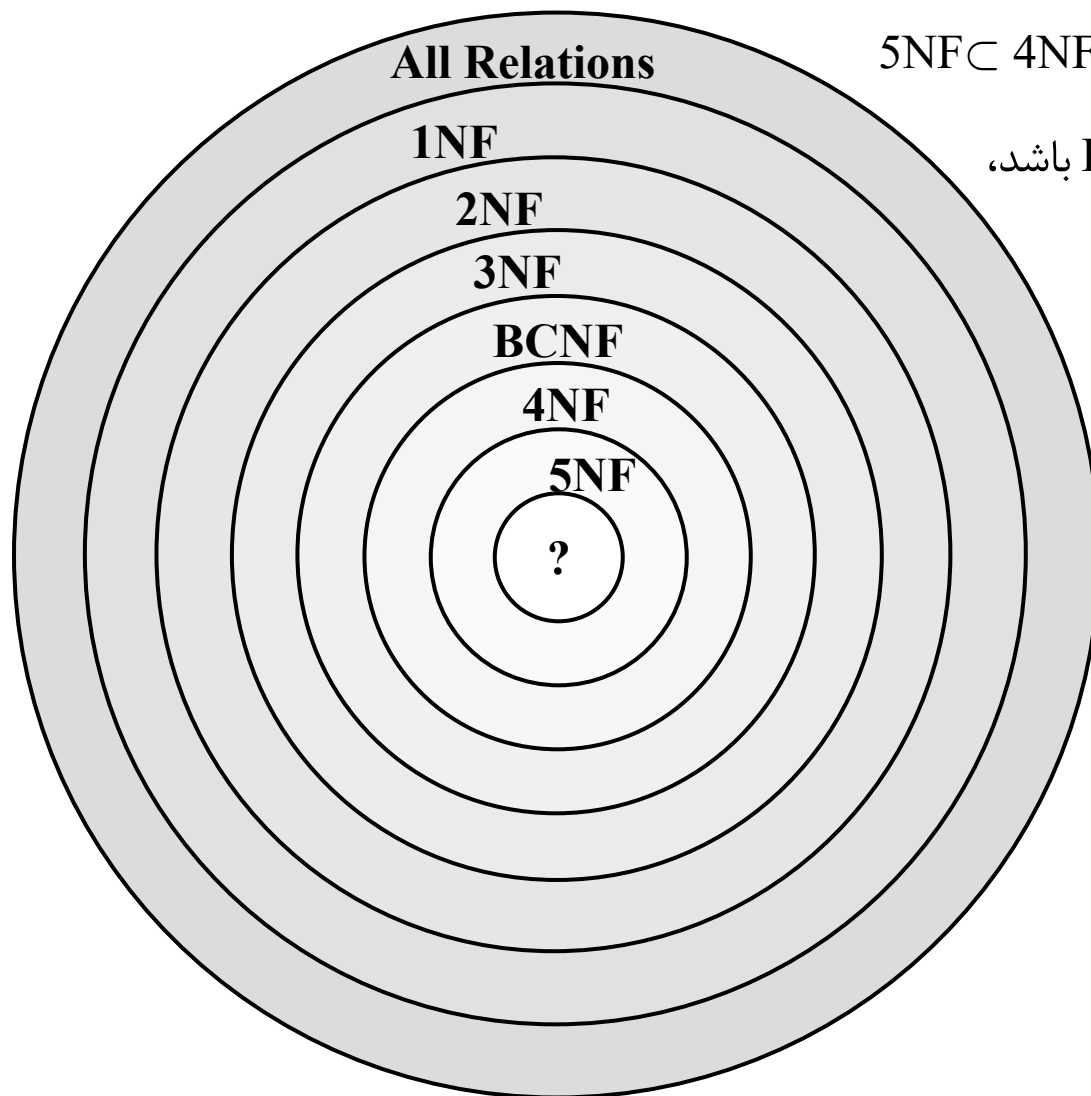
□ پس باید رابطه را نرمال‌تر کرد.



نرمال بودن رابطه (نرمالیتی)، فرم‌ها (صورت‌ها/ سطوح/ درجات) [NF: Normal Forms] مختلفی دارد.

فرم‌های نرمال:





$5NF \subset 4NF \subset BCNF \subset 3NF \subset 2NF \subset 1NF$ ☐

یعنی به طور مثال، رابطه‌ای که BCNF باشد، ☐

3NF هم هست.



□ برای بررسی فرم‌های نرمال، نیاز به مفاهیمی داریم از تئوری وابستگی (Dependency Theory).

□ مفاهیمی از تئوری وابستگی:

□ وابستگی تابعی (Functional Dependency)

□ وابستگی تابعی کامل [تام] (Fully Functional Dependency)

□ وابستگی تابعی با واسطه (Transitive Functional Dependency)



وابستگی تابعی (FD): صفت R.B به صفت R.A وابستگی تابعی دارد اگر و فقط اگر به ازای یک مقدار از A یک مقدار از B متناظر باشد. به عبارت دیگر اگر t_1 و t_2 دو تاپل از R باشند، در این صورت:

$$\text{IF } t_1.A = t_2.A \text{ THEN } t_1.B = t_2.B$$



با فرض اینکه کل تاپل‌های رابطه به صورت زیر باشد، آیا داریم:

R (A, B, C)

a_1 b_1 c_1

a_1 b_1 c_2

a_2 b_2 c_2

a_3 b_3 c_3

a_4 b_2 c_3

$$a_1 \rightarrow b_1$$

$$a_1 \begin{cases} c_1 \\ c_2 \end{cases}$$

$A \rightarrow B$ ؟ بله

$A \rightarrow C$ ؟ خیر

$B \rightarrow A$ ؟ خیر

$B \rightarrow C$ ؟ خیر



نکات: 

(۱) صفات طرفین FD می توانند ساده یا مرکب باشند.

(۲) اگر $A \rightarrow B$ ، لزوماً نداریم: $B \rightarrow A$.

(۳) اگر $B \subseteq A$ ، به $A \rightarrow B$ ، FD نامهم یا بدیهی (Trivial) گوییم.

(۴) اگر K در رابطه R ، SK یا CK باشد و $G \subseteq H_R$ آنگاه داریم: $K \rightarrow G$.

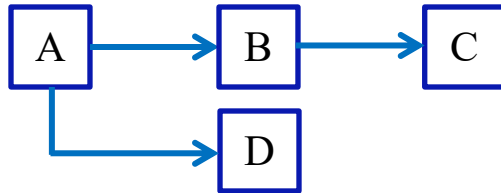


(۵) نمایش FDهای رابطه R به روشهای مختلف:

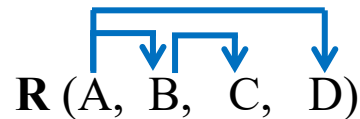
- به صورت یک مجموعه:

$$F = \{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$$

- با نمودار FDها:



- روی خود عنوان رابطه با استفاده از فلش‌هایی:





(۶) **تفسیر FD:** هر FD نمایشگر یک قاعده معنایی از محیط است: نوعی قاعده جامعیتی (که باید به نحوی به سیستم داده شود. خواهیم دید که در بحث طراحی، از طریق طراحی خوب به سیستم می‌دهیم).

□ **تمرین:** در رابطه $R(X, Y, Z)$ ، یک اظهار بنویسید که قاعده معنایی $X \rightarrow Y$ را پیاده‌سازی نماید.

(به طور مثال می‌توان از EXISTS استفاده کرد)

CREATE ASSERTION XTOYFD

CHECK (NOT EXISTS (SELECT X FROM R GROUP BY X HAVING MAX(Y) != MIN(Y)))

CONSTRAINT XTOYFD FORALL R1 (FORALL R2 IF R1.X=R2.X THEN R1.Y=R2.Y) حساب رابطه‌ای:

$STID \rightarrow STJ$: یک دانشجو فقط می‌تواند در یک رشته تحصیل کند.



$STJ \rightarrow STD$: یک رشته فقط در یک دانشکده ارائه می‌شود.

$STID \rightarrow STD$: یک دانشجو فقط در یک دانشکده تحصیل می‌کند.



قواعد استنتاج آرمسترانگ □

- 1- if $B \subseteq A$ then $A \rightarrow B \Rightarrow A \rightarrow A$ (قاعده انعکاسی)
- 2- if $A \rightarrow B$ and $B \rightarrow C$ then $A \rightarrow C$ (قاعده تعدی یا تراگذاری)
- 3- if $A \rightarrow B$ then $(A, C) \rightarrow (B, C)$ (قاعده افزایش)
- 4- if $A \rightarrow (B, C)$ then $A \rightarrow B$ and $A \rightarrow C$ (قاعده تجزیه)
- 5- if $A \rightarrow B$ and $C \rightarrow D$ then $(A, C) \rightarrow (B, D)$ (قاعده ترکیب)
- 6- if $A \rightarrow B$ and $A \rightarrow C$ then $A \rightarrow (B, C)$ (قاعده اجتماع)
- 7- if $A \rightarrow B$ and $(B, C) \rightarrow D$ then $(A, C) \rightarrow D$ (قاعده شبه تعدی)



□ سه قاعده اول **درست** و **کامل** هستند، بدین معنا که با داشتن یک مجموعه از وابستگی‌های تابعی F ،

تمام وابستگی‌های تابعی منطقاً قابل استنتاج از F ، با همین سه قاعده به دست می‌آیند و هیچ

وابستگی تابعی دیگر (که از F قابل استنتاج نباشد) نیز به دست نمی‌آید.

□ **توجه:** درستی سه قاعده اول به آسانی قابل اثبات است و قواعد دیگر از روی همانها اثبات می‌شوند.



وابستگی تابعی – قواعد آرمسترانگ (ادامه)

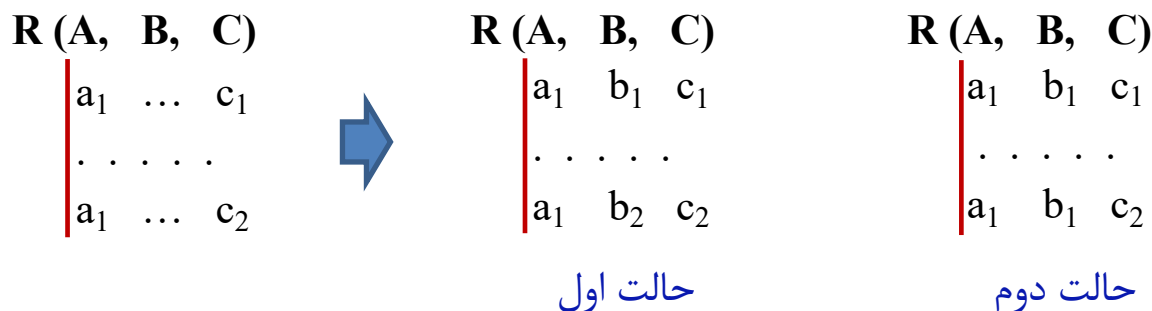
۱۵

بخش دهم: نرمال ترسازی پایگاه داده

تمرین: قاعده ۲ را اثبات کنید (با استفاده از برهان خلف).

اثبات: فرض خلف: گیریم که $A \not\rightarrow C$. در این صورت در رابطه R در حداقل دو تاپل، به ازای یک مقدار A ، دو مقدار متمایز از C داریم.

اما به ازای دو مقدار متمایز C ، مقدار B ممکن است دو مقدار متمایز با یک مقدار باشد.



در حالت اول، فرض $A \rightarrow B$ و در حالت دوم، فرض $B \rightarrow C$ نقض می شود. پس فرض خلف باطل است و حکم برقرار است.



□ کاربردهای قواعد آرمسترانگ

۱- محاسبه بستار صفت A : A^+

مجموعه تمام صفاتی که با A ، وابستگی تابعی دارند.

نکته: اگر $A \Leftarrow A^+ = H_R$ سوپرکلید (الگوریتم تشخیص سوپرکلید و نه کلید کاندید)

۲- محاسبه بستار مجموعه وابستگی‌های تابعی یک رابطه: F^+

مجموعه تمام FDهایی که از F منطقاً استنتاج می‌شوند:

$$F = \{A \rightarrow B, B \rightarrow C\} \Rightarrow F^+ = \{A \rightarrow B, B \rightarrow C, A \rightarrow C, (A, C) \rightarrow (B, C), \dots\}$$



□ کاربردهای مهم F^+ :

۱- تشخیص معادل بودن دو مجموعه از FDهای رابطه‌ای R: به طور نمونه F و G

□ شرط معادل بودن: $F^+ = G^+$

هر FD که از F به دست آید، از G هم به دست می‌آید.

۲- تشخیص FD افزونه

□ ضابطه تشخیص: وابستگی تابعی $f \in F$ را افزونه گوییم، هرگاه: $(F-f)^+ = F^+$

□ یعنی بود و نبود f در محاسبه F^+ تاثیری نداشته باشد.



۳- محاسبه مجموعه کاهش ناپذیر FD های یک رابطه

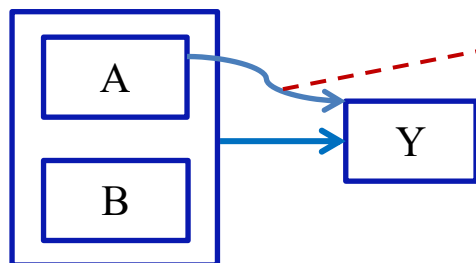
سه شرط دارد:

۱- هیچ FD در آن افزونه نباشد.

۲- سمت راست هر FD، صفت ساده باشد.

۳- سمت چپ هر FD، خود کاهش ناپذیر باشد: در وابستگی تابعی $X \rightarrow Y$ ، X را کاهش ناپذیر (و وابستگی $X \rightarrow Y$ را **کامل**) گوئیم، هرگاه Y با هیچ زیرمجموعه از X (غیر از خود X)، FD نداشته باشد. در غیر اینصورت X را کاهش پذیر گوئیم و وابستگی $X \rightarrow Y$ را **ناکامل** گوئیم.

اگر وجود داشته باشد، آنگاه X کاهش پذیر و $X \rightarrow Y$ یک FD ناکامل است.



$$\left\{ \begin{array}{l} (A, B) \rightarrow Y \\ A \rightarrow Y \end{array} \right. \Rightarrow \text{FD ناکامل}$$



□ **تمرین:** اگر یک FD کامل به صورت $A \rightarrow Y$ داشته باشیم، آنگاه FD ناکامل $(A,B) \rightarrow Y$ از آن قابل استنتاج است.

□ اثبات: با استفاده از قاعده افزایش از $A \rightarrow Y$ نتیجه می گیریم $(A,B) \rightarrow (Y,B)$

با استفاده از قاعده تجزیه داریم: $(A,B) \rightarrow B$ که یک FD بدیهی است و $(A,B) \rightarrow Y$ که همان حکم است.

مجموعه کاهش ناپذیر چه کاربردی دارد؟



□ **تعریف:** وابستگی تابعی با واسطه (TFD): اگر $A \rightarrow B$ ، $B \rightarrow C$ و $B \not\rightarrow A$ ، می گوئیم C با A، FD با واسطه از طریق B دارد.

اگر $B \rightarrow A$ هم برقرار باشد، آنگاه آن FD با واسطه، بدیهی (نامهم) است.



□ **توجه:** در سه فرم کلاسیک کادی، فقط با مفهوم کلید اصلی (PK) کار می‌کنیم و نه هر CK.

□ **تعریف:** 1NF: رابطه R در 1NF است اگر و فقط اگر تمام صفات آن تک‌مقداری باشد.

□ این تعریف می‌گوید هر رابطه نرمال در 1NF است.

□ **تعریف:** 2NF: رابطه R در 2NF است اگر و فقط اگر در 1NF باشد و هر صفت ناکلید (که خود PK یا CK

نباشد و جزء PK یا CK هم نباشد) در آن، با کلید اصلی رابطه، FD کامل داشته باشد.

□ به بیان دیگر در این رابطه FD ناکامل با کلید اصلی نداشته باشیم.

□ الگوریتم تبدیل 1NF به 2NF: حذف FDهای ناکامل از طریق تجزیه عمودی رابطه به طور مناسب.

□ **تعریف:** 3NF: رابطه R در 3NF است اگر و فقط اگر در 2NF باشد و هر صفت ناکلید با کلید اصلی رابطه، فقط

FD بی‌واسطه داشته باشد (FD با واسطه نداشته باشد).

□ الگوریتم تبدیل 2NF به 3NF: حذف FDهای با واسطه.



مثالی قید می‌کنیم و در آن تا 3NF پیش می‌رویم.



در حالت کلی، تمام صفات دانشجو، درس و انتخاب در یک رابطه می‌توانند باشند.

قواعد محیط:

R (STID, COID, STJ, STD, GR)

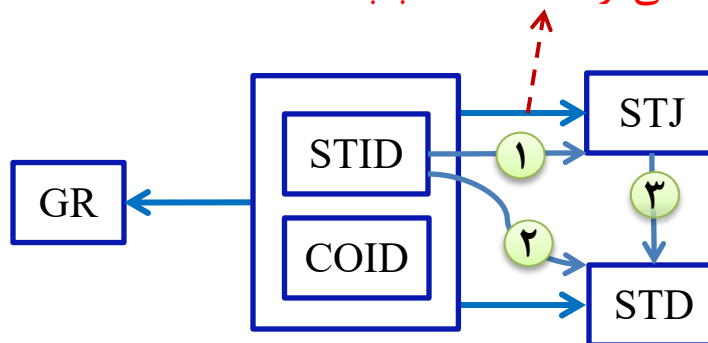
777	CO1	Phys	D11	19
777	CO2	Phys	D11	16
777	CO3	Phys	D11	11
888	CO1	Math	D12	16
888	CO2	Math	D12	18
444	CO1	Math	D12	13
555	CO1	Phys	D11	14
555	CO2	Phys	D11	12

۱- یک دانشجو در یک رشته تحصیل می‌کند.

۲- یک دانشجو در یک دانشکده تحصیل می‌کند.

۳- یک رشته در یک دانشکده ارائه می‌شود.

FDهای ناشی از PK (سمت چپ PK)





☐ رابطه R در 1NF است (چون همه صفات تک مقداری هستند) ولی آنومالی دارد و باید نرمال‌تر شود.

☐ آنومالی‌های رابطه R:

۱- در درج:

درج کن این فقره اطلاع درمورد یک دانشجو را: <'666', 'chem', 'D16'>

درج ناممکن: تا ندانیم حداقل یک درسی که گرفته شده چیست.

۲- در حذف:

فرض می‌کنیم '444' در این لحظه فقط همین تک درس را داشته باشد.

حذف کن فقط این اطلاع را: <'444', 'CO1', 13>

حذف انجام می‌شود اما اطلاع ناخواسته هم حذف می‌شود (اطلاعات دانشجو هم حذف می‌شود).

۳- در بهنگام‌سازی:

تغییر رشته تحصیلی دانشجو با شماره 777 به Chem.

برای انجام آن فزونکاری داریم؛ بهنگام‌سازی منتشرشونده (Propagating Update).



□ دلیل آنومالی‌های رابطه R:

□ از دیدگاه عملی: پدیده اختلاط اطلاعات، یعنی اطلاعات در مورد خود موجودیت دانشجو با اطلاعات در مورد انتخاب درس مخلوط شده است.

□ از دیدگاه تئوری: وجود FDهای ناکامل

$$\left\{ \begin{array}{l} (STID, COID) \rightarrow STJ \\ STID \rightarrow STJ \end{array} \right.$$

$$\left\{ \begin{array}{l} (STID, COID) \rightarrow STD \\ STID \rightarrow STD \end{array} \right.$$

□ این FDهای ناکامل باید از بین بروند. برای این منظور رابطه R را باید چنان تجزیه عمودی کنیم که در رابطه‌های حاصل، FD ناکامل نباشد.

□ برای این کار از عملگر **پرتو** استفاده می‌کنیم. پرتوی که منجر به یک **تجزیه خوب** شود.

$\Pi_{\langle \text{STID}, \text{COID}, \text{GR} \rangle}(\text{R})$



SCG (STID, COID, GR)

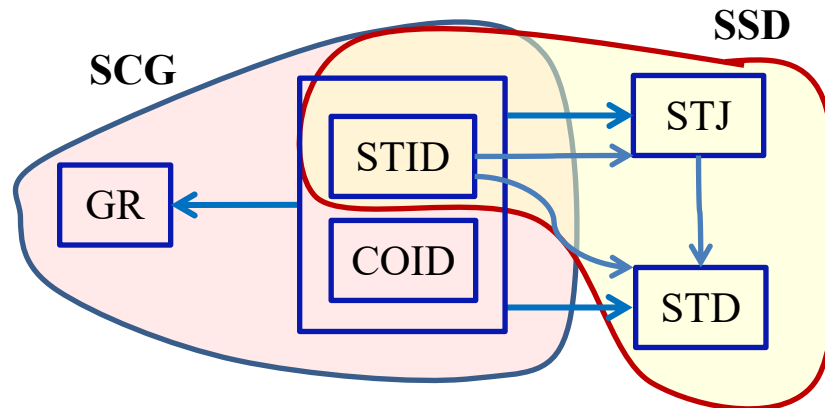
777	CO1	19
777	CO2	16
777	CO3	11
888	CO1	16
888	CO2	18
444	CO1	13
555	CO1	14
555	CO2	12

$\Pi_{\langle \text{STID}, \text{STJ}, \text{STD} \rangle}(\text{R})$

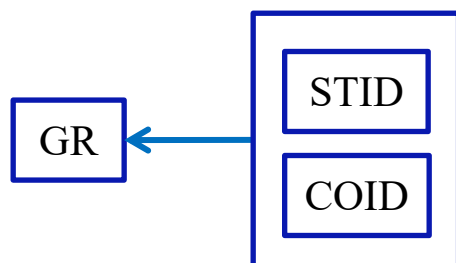


SSD (STID, STJ, STD)

777	Phys	D11
888	Math	D12
444	Math	D12
555	Phys	D11



SCG



رابطه‌های جدید آنومالی‌های R را ندارند: □

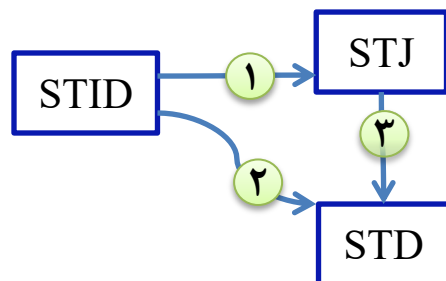
۱- درج کن: <'666', 'chem', 'D16'>

بدون مشکل در SSD درج می‌شود.

۲- حذف کن: <'444', 'CO1', 13>

بدون مشکل از SCG حذف می‌شود.

SSD



۳- بهنگام‌سازی کن: تغییر رشته دانشجوی 777 را به Chem

بدون مشکل در SSD بروز می‌شود.



□ در طراحی جدید، FDهای ناکامل از بین رفتند. بنابراین SSD و SCG، 2NF هستند.

□ **تاکید:** رابطه R، 2NF است هرگاه اولاً در 1NF باشد و ثانیاً هر صفت ناکلید با کلید اصلی، FD کامل

داشته باشد (رابطه، FD ناکامل نداشته باشد).

□ **تمرین:** بررسی شود که آیا در این تجزیه همه FDها محفوظ می‌مانند؟

□ **نکته:** باید توجه کنیم که در تجزیه، FDای از دست نرود، چون هر FD یک قاعده جامعیت در محیط است.

□ توجه داشته باشید که در این تجزیه هیچ اطلاعی از دست نمی‌رود. یعنی اگر کاربر رابطه اصلی را به هر

دلیلی بخواهد، با پیوند دو رابطه جدید به دست می‌آید.

$$R = SCG \bowtie SSD$$



□ آیا رابطه‌های جدید (SSD و SCG) آنومالی ندارند؟

□ آنومالی‌های SSD:

۱- در درج:

اطلاع: «رشته IT در دانشکده D20 ارائه می‌شود.» به دلیل FD شماره ۳، این اطلاع منطقاً باید قابل درج باشد، اما درج ناممکن است. چون کلید ندارد، باید حداقل یک دانشجوی این رشته را بشناسیم.

۲- در حذف:

حذف کن «Chem»، «666» و با فرض اینکه تنها یک دانشجو در رشته Chem ثبت شده است. حذف انجام می‌شود ولی اطلاع «رشته شیمی در D16 ارائه می‌شود»، ناخواسته حذف می‌شود.

۳- در بهنگام‌سازی:

«شماره دانشکده رشته فیزیک را عوض کنید». به تعداد تمام دانشجویان این رشته باید بهنگام‌سازی شود.

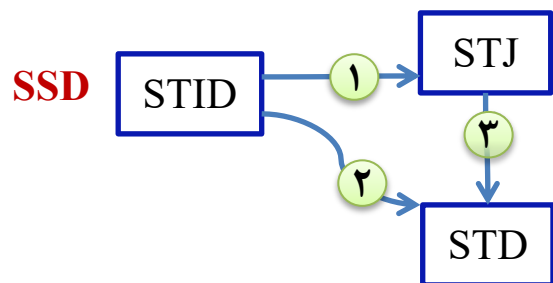
SSD باید نرمال تر شود.





□ دلیل آنومالی‌های SSD:

□ دلیل آنومالی‌های SSD، وجود FD با واسطه بین صفت ناکلید با کلید اصلی است (به دلیل FD شماره ۳).



□ این FD باید از بین برود.

□ فرض کنید SSD را به صورت زیر تجزیه کنیم:

$SJ(\underline{STID}, \underline{STJ})$ و $SD(\underline{STJ}, \underline{STD})$

777	Phys
888	Math
444	Math
555	Phys

Phys	D11
Math	D12

□ افزودن کمی کم شد!

□ تمرین: بررسی شود که رابطه‌های جدید آنومالی‌های SSD را ندارند.

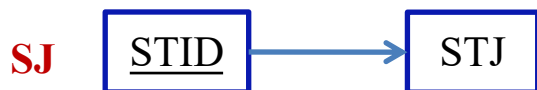


فرم‌های نرمال کلاسیک کادی (ادامه)

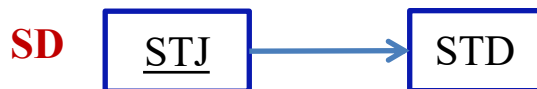
۲۹

بخش دهم: نرمال ترسازي پایگاه داده

☐ این رابطه‌ها در 3NF هستند.



☐ اولاً در 2NF هستند.



☐ ثانیاً با FD واسطه نداریم.

☐ **تمرین:** بررسی شود که در این تجزیه هیچ اطلاعی از دست نمی‌رود و FDها هم حفظ می‌شوند.

☐ **تاکید:** رابطه R در 3NF است اگر و فقط اگر اولاً در 2NF باشد و ثانیاً هر صفت ناکلید با کلید اصلی FD

بی‌واسطه داشته باشد (تمام FDها مستقیماً ناشی از PK باشد).

☐ **نتیجه:** FDهای ناکامل و باواسطه مزاحم هستند و باید از بین بروند.

☐ در عمل رابطه‌ها باید حداقل تا 3NF نرمال شوند و خواهیم دید حتی‌الامکان در BCNF یا بیشتر باشند.

☐ در رابطه 3NF داریم که «یک بوده (واقعیت): یک رابطه» و یا «یک شیء: یک رابطه».



□ در حالت کلی اگر R_1, R_2, \dots, R_n پرتوهای دلخواه از R باشند، به شرط عدم وجود هیچمقدار در صفات پیوند داریم (ممکن است تاپل‌های افزونه بروز کند):

$$R \subseteq R_1 \bowtie R_2 \bowtie \dots \bowtie R_n$$

□ **تجزیه بی حذف:** شرطش این است که **در صفات پیوند هیچمقدار (Null Value)** نداشته باشیم.

□ اگر در صفات پیوند هیچمقدار داشته باشیم، چه پیش می‌آید؟

$$T(\underline{A}, B, C, D, E) \Rightarrow T_1(A, B) \quad T_2(B, C, D, E)$$

تاپل‌هایی در پیوند از دست می‌روند. به این تاپل‌ها، تاپل‌های آونگان [معلق] (Dangling) می‌گوییم.

□ در مباحث نرمال‌ترسازی معمولاً فرض بر این است که **صفت (صفات) پیوند، هیچمقدار ندارند.**



□ تجزیه خوب (Nonloss/Lossness Decomposition)

۱- بی حشو: در پیوند پرتوها، تاپل حشو [افزونه] بروز نکند.

۲- حافظ FD ها: هیچ FD ای در اثر تجزیه از دست نرود و همه FD های رابطه اصلی حفظ شوند.

۳- بی حذف: در پیوند پرتوها هیچ تاپلی حذف نشود (صفت یا صفات پیوند هیچمقدار نباشند).

۴- حافظ صفات: $\bigcup_{i \in \{1, \dots, n\}} H_{R_i} = H_R$

پیش فرض یا بدیهی

□ در بیشتر متون کلاسیک، **تجزیه بی حشو** تحت عنوان **تجزیه بی کاست یا بی گمشدگی**

(Nonloss/Lossless Decomposition) مطرح شده است که به همراه خاصیت حفظ وابستگی های تابعی،

تجزیه خوب را شکل می دهد (دو ویژگی دیگر تجزیه خوب را پیش فرض تجزیه خوب می دانیم).

□ در واقع تاپلهای افزونه باعث از دست رفتن بخشی از اطلاعات می شوند.



□ قضیه ريسانن (Rissanen):

□ رابطه R به دو پرتوش $(R_1$ و $R_2)$ **تجزیه خوب** می شود، اگر R_1 و R_2 از یکدیگر مستقل باشند.

□ R_1 و R_2 **مستقل** از یکدیگرند اگر و فقط اگر:

- صفت مشترک، حداقل در یکی از آنها CK باشد \Leftarrow بی حشو بودن

- تمام FD های رابطه اصلی یا در مجموعه FD های R_1 و R_2 وجود داشته باشند یا از آنها منطقی

استنتاج شوند \Leftarrow حافظ FD ها

□ **نکته:** بر اساس ضوابط ريسانن، اگر در رابطه $R(A, B, C)$ وابستگی های $A \rightarrow B$ ، $A \rightarrow C$ و $B \rightarrow C$ برقرار

باشد، در این صورت تجزیه خوب چنین است: $R_1(\underline{A}, B)$ و $R_2(\underline{B}, C)$.

□ در اینجا صفت مشترک B در رابطه دوم کلید کاندید است، چون همه صفات به آن وابستگی تابعی

دارند و کاهش پذیر هم نیست.



مثال: رابطه SSD را در نظر می‌گیریم. این رابطه به سه شکل به پرتوهای دوگانی قابل تجزیه است. ☐

- I SS (STID, STJ) SD (STJ, STD)
- II SS (STID, STJ) SD (STID, STD)
- III SS(STID, STD) SJ (STJ, STD)

☐ تجزیه I خوب است، چون هر دو شرط ریساین را دارد.

$$\left. \begin{array}{l} \text{STID} \rightarrow \text{STJ} \\ \text{STJ} \rightarrow \text{STD} \end{array} \right\} \Rightarrow \text{STID} \rightarrow \text{STD}$$

☐ تجزیه II خوب نیست، چون FD از دست می‌دهد.

☐ تجزیه III خوب نیست، چون FD از دست می‌دهد.



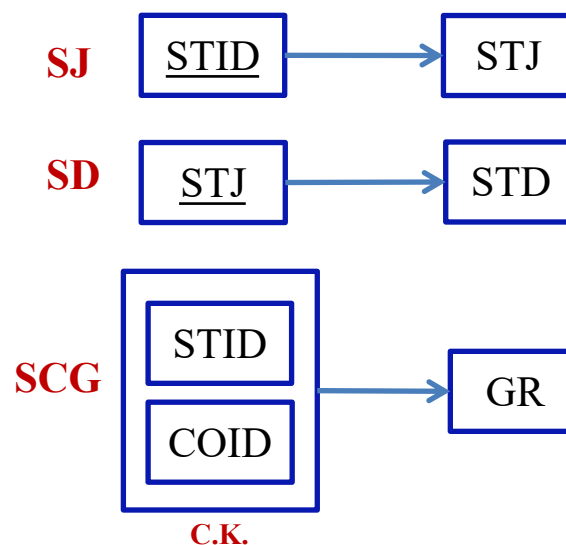
□ **اصطلاح:** در وابستگی تابعی $A \rightarrow B$ (A Determines B) به A دترمینان گویند.

□ **تعریف:** رابطه R در BCNF است اگر و فقط اگر در آن دترمینان هر FD مهم و کاهش ناپذیر، CK باشد.

□ چون رابطه می تواند بیش از یک CK داشته باشد، BCNF از 3NF قوی تر است.

مثال: رابطه های زیر در BCNF هستند.

SCGJD { SCG(SID, COID, GR)
SJ (STID, STJ)
SD (STJ, STD)





□ BCNF از 3NF قوی تر است. \Leftarrow رابطه می تواند در 3NF باشد، اما در BCNF نباشد.

□ **حالت I:** رابطه R فقط یک CK داشته باشد. \Leftarrow اگر R در 3NF باشد، در BCNF هم هست (مثال دیده شده).

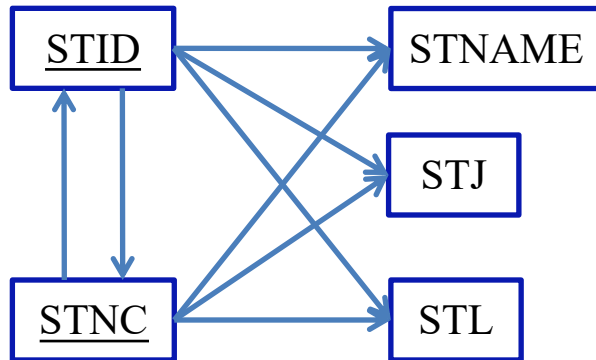
□ **حالت II:** رابطه R بیش از یک CK داشته باشد.

□ **II-1)** CKها مجزا باشند (صفت مشترک نداشته باشند). \Leftarrow اگر R در 3NF باشد، در BCNF هم هست.

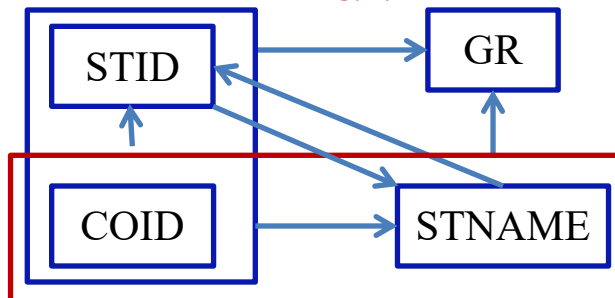
□ **II-2)** CKها هم پوشا باشند. \Leftarrow اگر R در 3NF باشد، لزوماً در BCNF نیست.



ST (STID, STNAME, STNC, STJ, STL, ...)
C.K. C.K.



SCNG (STID, COID, STNAME, GR)
C.K. C.K.



برای حالت II-1



دو دترمینان، هر دو هم CK هستند.

برای حالت II-2



(فرض: هیچ دو دانشجویی نام یکسان ندارند.)



□ کافی است یک دترمینان در رابطه پیدا کنیم که CK نباشد. \Leftarrow رابطه BCNF نیست.

□ پس در کدام فرم نرمال است؟

□ 1NF هست. چون صفت‌ها تک‌مقداری هستند.

□ 2NF هست. چون FD ناکامل نداریم. \Leftarrow هر صفت ناکلید با کلید اصلی FD ناکامل نداشته باشد.

\Leftarrow در اینجا STNAME صفت غیرکلید نیست، پس FD ناکامل نیست.

□ 3NF هست. چون FD با واسطه با کلید اصلی نداریم.

□ آیا این رابطه تجزیه می‌شود؟

$\left\{ \begin{array}{l} \text{SCG}(\underbrace{\text{STID}}_{\text{C.K.}}, \text{COID}, \text{GR}) \\ \text{SSN}(\underbrace{\text{STID}}_{\text{C.K.}}, \underbrace{\text{STNAME}}_{\text{C.K.}}) \end{array} \right. \Rightarrow \text{هر دو BCNF هستند.}$

□ آیا طرز دیگر هم می‌شود تجزیه کرد؟ بله، به جای STID در SCG، STNAME بگذاریم.



□ نشان دهید که این تجزیه خوب است؛ یعنی با پیوند پرتوها، رابطه اصلی به دست می‌آید و هیچ FD از دست نمی‌رود.

□ چه پدیده‌ای در اینجا دیده می‌شود؟ این رابطه اختلاط اطلاعات دارد! با این همه 3NF است.

SCNG (STID, COID, STNAME, GR)
C.K. C.K.

□ نکته: صرف وجود اختلاط اطلاعات ایجاب می‌کند که رابطه در فرم نرمال ضعیفی باشد.

□ تمرین: محیط دانشکده، قواعد معنایی:

۱- یک دانشجو یک درس را با یک استاد انتخاب می‌کند.

۲- یک استاد فقط یک درس تدریس می‌کند.

۳- یک درس توسط بیش از یک استاد ارائه می‌شود.



فرض می کنیم طراح رابطه زیر را طراحی کرده است.



این رابطه در کدام فرم نرمال است؟

ابتدا باید با استفاده از قواعد، CKها را مشخص کنیم. سپس نمودار FD را رسم کنیم.

آیا این رابطه، تجزیه خوب دارد؟

نکته: اگر رابطه مثلاً 3NF باشد و تجزیه خوب نداشته باشد، نباید تجزیه کنیم تا رابطه های حاصل

BCNF باشد.

رابطه فوق در 3NF است و از نکته فوق این نتیجه مهم به دست می آید که این رابطه تجزیه خوب ندارد.

نکته: رابطه تمام کلید حداقل BCNF است.



قضیه هیت (Heath): در رابطه $R(A, B, C)$ ، که در آن A ، B و C سه مجموعه از صفات هستند، اگر $A \rightarrow B$ (در F^+ باشد)، آنگاه تجزیه R به دو پرتو $R_1(A, B)$ و $R_2(A, C)$ **تجزیه بی کاست** (Nonloss) است.

دقت شود که برقراری شرایط **قضیه هیت**، یک تجزیه بی کاست (و نه لزوماً خوب که حافظ FD باشد) را تضمین می‌نماید اما برقراری شرایط **قضیه ريسانن**، یک تجزیه خوب را تضمین می‌نماید. واضح است که در قضیه ريسانن شرایط قضیه هیت نیز برقرار است. بیانی دیگر از **قضیه هیت** تحت عنوان **تست NJB** به صورت زیر است.

تست پیوند بی حشو برای تجزیه دودویی (NJB- Nonadditive Join Test for Binary Decompositions):

تجزیه دودویی $D = \{R_1, R_2\}$ از رابطه R خاصیت پیوند بی حشو دارد اگر و تنها اگر یکی از موارد زیر با توجه به مجموعه FD های F برقرار باشد:

- وابستگی تابعی $(R_1 - R_2) \rightarrow (R_1 \cap R_2)$ در F^+ باشد یا

- وابستگی تابعی $(R_2 - R_1) \rightarrow (R_1 \cap R_2)$ در F^+ باشد.



4NF: رابطه R در 4NF است اگر و فقط اگر در BCNF باشد و وابستگی چندمقداری (MVD) مهم در آن وجود نداشته باشد.

وابستگی چندمقداری (MVD): در رابطه $R(A, B, C)$ (رابطه با سه صفت یا سه مجموعه صفت)، صفت B با صفت A، MVD دارد $(A \twoheadrightarrow B)$ اگر و فقط اگر به ازای یک مقدار A، مجموعه‌ای از مقادیر B متناظر باشد.

[یعنی به ازای هر جفت مشخص از (A,C)، مجموعه مقادیر B فقط با تغییرات A تغییر کند.]

به فرم نرمال
 $R(A, B, C)$

a ₁	b ₁	c ₁
a ₁	b ₂	c ₁
a ₁	b ₃	c ₁
a ₁	b ₁	c ₂
a ₁	b ₂	c ₂
a ₁	b ₃	c ₂
a ₂	b ₁	c _i
a ₂	b ₇	c _i

به فرم غیرنرمال
 $R(A, B, C)$

a ₁	$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$	c ₁
a ₁	$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$	c ₂
a ₂	$\begin{bmatrix} b_1 \\ b_7 \end{bmatrix}$	c _i



$A \twoheadrightarrow B$



نکات: □

۱- اگر $B \subseteq A$ باشد، به $A \rightarrow \rightarrow B$ می‌گوییم MVD بدیهی [نامهم]

اگر $A \cup B = H_R$ باشد، به $A \rightarrow \rightarrow B$ می‌گوییم MVD بدیهی [نامهم]

۲- MVD در رابطه‌های با سه صفت [ساده یا مرکب] همیشه جفت است.

If $A \rightarrow \rightarrow B$ then $A \rightarrow \rightarrow (H - \{A, B\})$ یا $A \rightarrow \rightarrow C$

برای اثبات این نکته کافی است به جای یک جفت مقدار از (A, C) ، یک جفت (A, B) را بگیریم، آن مجموعه برای C تشکیل می‌شود.

۳- برای MVD هم قواعد آرمسترانگ وجود دارد که با قواعد مربوط به FDها متفاوت است.



استاد از دانشجو گزارش آزمایشگاه می‌گیرد.



رابطه غیرنرمال با صفت چندمقداری

NNPSR (PR#, ST#, RE#)

□ در این محیط یک قاعده معنایی خاص وجود دارد: یک استاد از هر یک از دانشجویان یک گروه، هر

یک از گزارش‌های یک مجموعه گزارش را می‌گیرد.

□ اگر این قاعده معنایی نباشد، این مجموعه‌ها شکل نمی‌گیرد.

NNPSR (PR#, ST#, RE#)

$PR_1 \begin{bmatrix} 777 \\ 888 \\ 444 \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}$

$PR_2 \begin{bmatrix} 777 \\ 666 \end{bmatrix} \begin{bmatrix} R_3 \end{bmatrix}$

...



رابطه غیرنرمال با صفت چندمقداری

NNCTX (C#, T#, B#)

c ₁	t ₁	b ₁
	t ₂	b ₂
	t ₃	
c ₂	t ₄	b ₃
		b ₅
	t ₂	b ₇

درس C توسط استاد T از روی کتاب مرجع B ارائه می‌شود.



پدیده MVD بیان فرمال صفت چندمقداری است. □

CTX (C#, T#, B#)

c ₁	t ₁	b ₁
c ₁	t ₂	b ₁
c ₁	t ₃	b ₁
c ₁	t ₁	b ₂
c ₁	t ₂	b ₂
c ₁	t ₃	b ₂
c ₂	t ₄	b ₃
c ₂	t ₂	b ₃
c ₂	t ₄	b ₅
c ₂	t ₂	b ₅
c ₂	t ₄	b ₇
c ₂	t ₂	b ₇

فرم نرمال شده این مثال، افزونگی زیادی دارد.



رابطه تمام کلید است؛ یعنی هیچ یک به تنهایی و □

هیچ ترکیب دوتایی آن CK نیست.

رابطه تمام کلید حداقل BCNF است. □

زیرا یک دترمینان دارد که آن هم CK است.



با این همه رابطه اخیر آنومالی دارد. ☐

☐ **در درج:** در درس c_1 ، کتاب b_8 نیز به عنوان مرجع درس ثبت شود.

نمی توانیم بگوییم چون کلید نداریم نمی توانیم درج کنیم. باید قواعد معنایی رعایت شود.

باید درج کنیم: $\langle c_1, t_1, b_8 \rangle$

$\langle c_1, t_2, b_8 \rangle$

$\langle c_1, t_3, b_8 \rangle$

یعنی عمل منطقاً تاپلی تبدیل شده به عمل مجموعه ای

☐ در حذف و بهنگام سازی هم به دلیل وجود افزونگی، آنومالی داریم.

☐ رابطه CTB باید تجزیه شود تا رابطه های حاصل 4NF شود.



□ دلیل آنومالی این رابطه، وجود پدیده MVD است.

$$\left\{ \begin{array}{l} C\# \rightarrow\rightarrow B\# \\ C\# \rightarrow\rightarrow T\# \end{array} \right.$$

□ پس CTB را باید چنان تجزیه کنیم که در رابطه‌های حاصل، MVD مهم وجود نداشته باشد.

□ برای این کار CTB را پرتوگیری می‌کنیم به نحوی که در عنوان هر پرتو، مبدأ MVD وجود داشته باشد.

CT (C#, T#)

c ₁	t ₁
c ₁	t ₂
c ₁	t ₃
c ₂	t ₄
c ₂	t ₂

CB (C#, B#)

c ₁	b ₁
c ₁	b ₂
c ₂	b ₃
c ₂	b ₅
c ₂	b ₇
c ₁	b ₈

درج به صورت عملاً تاپلی و نه مجموعه‌ای

□ رابطه‌های جدید آنومالی CTB را ندارند.

□ این دو رابطه جدید BCNF هستند، چون تمام کلید هستند. MVD مهم ندارند، پس 4NF هستند.

□ **تمرین:** نشان دهید با پیوند این دو رابطه، رابطه اصلی به دست می‌آید.



□ قضیه فاگین (Fagin): رابطه $R(A, B, C)$ به دو پرتوش $R_1(A, B)$ و $R_2(A, C)$ تجزیه بی کاست

(Nonloss) می شود اگر و فقط اگر $A \rightarrow \rightarrow B$.

□ قضیه فاگین (برای MVD) تعمیم قضیه هیث (برای FD) است.

□ آیا می توان گفت مفهوم MVD تعمیم مفهوم FD است؟ آیا می توان گفت FD حالت خاصی از MVD است؟

□ FD حالت خاصی از MVD است که در آن مجموعه مقادیر صفت وابسته، تک عنصری هستند.

□ همچنین این استنتاج منطقی را هم داریم:

If $A \rightarrow B$ then $A \rightarrow \rightarrow B$



❑ **نکته:** بحث 4NF از یک دیدگاه می تواند اصلاً موضوعیت نداشته باشد. زیرا رابطه‌ای که BCNF باشد و

MVD داشته باشد قطعاً صفت چندمقداری دارد و می دانیم در طراحی برای صفات چندمقداری، از همان

ابتدا می توان رابطه‌های جداگانه طراحی کرد.

❑ با این همه مفهوم MVD به عنوان بیان فرمال صفت چندمقداری قابل توجه است.

تعریف زاینولو از 3NF، BCNF، 4NF و ... مطالعه شود.





وابستگی پیوندی (JD): رابطه R وابستگی پیوندی به n پرتو R_1, R_2, \dots, R_n دارد اگر و فقط اگر R



حاصل پیوند بی حشو این n پرتو باشد.

$$R = [JD] * (R_1, R_2, \dots, R_n)$$

$$CTB = [JD] * (CT, CB)$$



JD را نامهم گوئیم هرگاه عنوان (Heading) یکی از R_i ها همان عنوان (Heading) رابطه R باشد. □

5NF [PJNF] – فرم نرمال پرتو پیوندی: رابطه R در 5NF است اگر و فقط اگر تمام JDهای آن ناشی



از CK باشد. \Leftarrow ناشی از CK بودن یعنی عنوان همه پرتوها، در همه JDها، سوپرکلید باشد.

□ رابطه CTB در 5NF نیست، چون $(C\#, T\#)$ و $(C\#, B\#)$ سوپرکلید رابطه CTB نیستند.



STUD (STID, STNAME, STJ, STL)



□ فرض می‌کنیم که 3NF هست و FD مزاحم نداریم.

$$\left\{ \begin{array}{l} \text{STN} (\underline{\text{STID}}, \text{STNAME}) \\ \text{SJL} (\underline{\text{STID}}, \text{STJ}, \text{STL}) \end{array} \right. \Rightarrow \text{STUD} = [\text{JD}] * (\text{STN}, \text{SJL}) \quad \text{JD به دو پرتو}$$

$$\left\{ \begin{array}{l} \text{STN} (\underline{\text{STID}}, \text{STNAME}) \\ \text{SJ} (\underline{\text{STID}}, \text{STJ}) \\ \text{SL} (\underline{\text{STID}}, \text{STL}) \end{array} \right. \Rightarrow \text{STUD} = [\text{JD}] * (\text{STN}, \text{SJ}, \text{SL}) \quad \text{JD به سه پرتو}$$

□ رابطه STUD در 5NF است. چون عنوان همه پرتوها در همه JDهای آن، سوپرکلید هستند (ناشی از کلید کاندید هستند).



نکته: اگر رابطه‌ای در 3NF باشد و تمام CKهای آن ساده باشند، آن رابطه در 5NF است.

مثال PCD رابطه‌ای است که در 4NF است ولی در 5NF نیست. JD به دو پرتو ندارد، بلکه به سه پرتوش JD دارد، نه هیچکدام سوپرکلید نیستند. این محدودیت دائمی و مستقل از زمان است که از محدودیت‌های محیط عملیاتی نشأت گرفته و همواره در بدنه رابطه برقرار است.

PCD (PR#, CO#, D#)

PR ₁	co ₁	d ₂
PR ₁	co ₂	d ₁
PR ₂	co ₁	d ₁
PR ₁	co ₁	d ₁

محدودیت موجود: هرگاه استاد p درس c را ارایه کند و درس c در دانشکده d ارایه شده باشد و استاد p در دانشکده d حداقل یک درس ارایه داده باشد، آنگاه استاد p درس c را نیز در دانشکده d ارایه می‌نماید.

رابطه SPJ تمام کلید است. \Leftarrow حداقل BCNF

MVD ندارد. \Leftarrow 4NF است.

SPJ (S#, P#, J#)

S ₁	P ₁	J ₁
S ₁	P ₁	J ₂
S ₁	P ₂	J ₁
S ₂	P ₁	J ₁

محدودیت موجود: هرگاه تولیدکننده S قطعه p را تولید کند و قطعه p در پروژه j استفاده شود و تولیدکننده S حداقل یک قطعه در پروژه j تولید کرده باشد، آنگاه تولیدکننده S قطعه p را در پروژه j تولید کرده است.



بخش دهم: نرمال ترسازی پایگاه داده

فرض می کنیم بخواهیم این رابطه را تجزیه کنیم: ☐

SP (S#, P#)

S ₁	P ₁
S ₁	P ₂
S ₂	P ₁

PJ (P#, J#)

P ₁	J ₁
P ₁	J ₂
P ₂	J ₁



SPJ' (S#, P#, J#)

S ₁	P ₁	J ₁
S ₁	P ₁	J ₂
S ₁	P ₂	J ₁
S ₂	P ₁	J ₂
S ₂	P ₁	J ₁

تاپل حشو

SJ (S#, J#)

S ₁	J ₂
S ₁	J ₁
S ₂	J ₁

این رابطه JD به دو پرتوش ندارد. ☐

یک پرتو دیگر هم می گیریم: ☐



SPJ (S#, P#, J#)

۱	S ₁	P ₁	J ₁
۲	S ₁	P ₁	J ₂
۳	S ₁	P ₂	J ₁
۴	S ₂	P ₁	J ₁



$$SPJ = [JD] * (SP, PJ, SJ)$$

□ پس SPJ، JD دارد به سه پرتوش و نه کمتر:

و 5NF نیست چون عنوان (Heading) پرتوهایش سوپرکلید نیست.

□ در این مثال از سه فقره اطلاع دو موجودیتی، باید یک اطلاع سه موجودیتی را استنتاج کنیم، چرا که این

یک محدودیت جامعیتی حاکم بر محیط است (وجود وابستگی پیوندی).

□ توجه داشته باشید که در حالت کلی چنین استنتاجی درست نیست و پدیده دام پیوندی حلقه‌ای بروز

می‌کند، ولی در اینجا به دلیل وجود وابستگی پیوندی، چنین مشکلی بروز نمی‌کند.



نکته: در این رابطه یک محدودیت بسیار نادر، موسوم به محدودیت با **ماهیت چرخشی (CC)** وجود دارد.

□ با وجود تاپل‌های دوم تا چهارم در رابطه SPJ باید تاپل (S_1, P_1, J_1) نیز وجود داشته باشد.

□ این محدودیت ناشی از وجود (S_1, P_1) در تاپل دوم، (S_1, J_1) در تاپل سوم و (P_1, J_1) در تاپل چهارم است.

□ در واقع مقدار هر یک از سه صفت در سه تاپل از چهار تاپل رابطه SPJ یکسان است و در هر یک از سه پرتو دوتایی، یک صفت مشترک با دو پرتو دیگر وجود دارد.

□ اگر یک رابطه CC داشته باشد در فرم نرمال 5NF نیست.

□ برای تشخیص این محدودیت در رابطه درجه n دوتست انجام می‌دهیم:

۱- تعداد تاپل‌ها: $n+1$

۲- مقدار هر صفت، در n تاپل یکسان باشد.



در رابطه R هر ترکیب دوتایی CK است. لذا در فرم نرمال 5NF است زیرا:



R (A, B, C)

a ₁	b ₁	c ₂
a ₁	b ₂	c ₁
a ₂	b ₁	c ₁

□ سه دترمینان دارد که هر سه CK هستند. \Leftarrow BCNF است.

□ MVD ندارد. \Leftarrow 4NF است.

□ CC ندارد و همه JDهای آن ناشی از کلید کاندید هستند. \Leftarrow 5NF است.



فرم نرمال 6NF – جهت مطالعه بیشتر

۵۶

بخش دهم: نرمال ترسازی پایگاه داده



رابطه R در 6NF است هر گاه اصلاً JD مهم نداشته باشد.

□ **نکته:** در رابطه درجه n، اگر غیر از کلید فقط یک صفت دیگر داشته باشد، در 6NF است.



به طور مثال رابطه SPJ که 5NF نبود را به سه رابطه SP، SJ و PJ تجزیه می کنیم.
این سه رابطه در فرم نرمال 5NF و 6NF هستند.



فرم نرمال DKNF چیست؟



□ تئوری نرمال‌ترسازی به عنوان ابزار طراحی RDB، مزایا و معایبی دارد.

□ مزایای تئوری نرمال‌ترسازی:

۱- ارائه یک طراحی واضح از خُردجهان واقع (Clean Design)؛ یعنی با کمترین اختلاط اطلاعات.

یعنی در واقع رعایت یک اصل در عمل (one fact : one table).

۲- کاهش بعض افزونگی‌ها؛ آن افزونگی‌هایی که با پرتوگیری از بین می‌روند (کاهش می‌یابد).

۳- کاهش بعض آنومالی‌ها [ناشی از اختلاط اطلاعات].

۴- بعض قواعد جامعیت را اعمال می‌کنیم (ناشی از وابستگی بین صفات).

□ این تئوری به طراح کمک می‌کند تا تصمیم بگیرد چند رابطه داشته باشد و هر رابطه عنوانش چه باشد و

کلیدش چه باشد.



❑ معایب تئوری نرمال‌ترسازی:

- ۱- فزون‌کاری در بازیابی (اگر کاربر به هر دلیلی رابطه اصلی را بخواهد، عمل پیوند (Join) باید انجام شود که در حجم بالای داده، سربار زیادی دارد).
به دلیل همین عیب، گاه در عمل لازم است غیرنرمال‌سازی (Denormalization) انجام دهیم.
یعنی تبدیل حداقل دو رابطه $(i+1)NF$ به یک رابطه iNF .
- ۲- فرآیند نرمال‌ترسازی زمان‌گیر است به ویژه اگر مجموعه صفات محیط بزرگ باشد و نمودار FDها گسترده باشد.
- ۳- مبتنی است بر یک فرض نه چندان واقع‌بینانه [فرض: در آغاز مجموعه‌ای از صفات داریم در یک مجموعه Universal، آنگاه با روش سنتز صفات (دسته‌بندی صفات) به تعدادی رابطه می‌رسیم.] در حالیکه در عمل ابتدا روش بالا به پایین و رسیدن به تعدادی رابطه با درجه متعارف، آنگاه استفاده از ایده‌های این تئوری برای تست نرمالیتی (اول تست $3NF$ ، بعد BCNF و $5NF$).



۴- همه وابستگی‌های بین صفات دیده نشده‌اند؛ مثلاً وابستگی شمول دیده نشده است.

۵- ایجاد میزانی افزونگی؛ چون اگر بخواهیم تجزیه خوبی داشته باشیم، یا CK باید در همه پرتوها تکرار شود یا پیوندهای CK-FK وجود داشته باشد!

۶- استفاده محدود از عملگرهای جبر رابطه‌ای. تجزیه ← پرتو بازسازی ← پیوند
حال آنکه در عمل گاه لازم است رابطه را تجزیه افقی کنیم:

$$ST_1 = \sigma_{STJ='Phys'}(STUD)$$

$$ST_2 = \sigma_{STJ='IT'}(STUD)$$

...

$$ST_n = \sigma_{STJ='Comp'}(STUD)$$

$$STUD = \bigcup_{i=1}^n (ST_i)$$



☐ به رابطه‌های ناشی از تجزیه افقی می‌گوییم:

فرم نرمال‌گزینش اجتماع (تحدید اجتماع) RUNF (Restriction Union Normal Form)

☐ RUNF لزوماً در امتداد فرم‌های نرمال نیست. به موازات آنها مطرح است. یعنی ممکن است رابطه 3NF باشد، تجزیه افقی کنیم و باز هم 3NF باشد.

☐ در چه شرایطی رابطه حاصل از تجزیه افقی از خود رابطه نرمال‌تر است؟



پرسش و پاسخ ...

amini@sharif.edu