

به نام آنکه جان را فکرت آموخت



## بخش هفتم: طراحی پایگاه داده رابطه‌ای

مرتضی امینی

نیمسال اول ۹۲-۹۳

(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

کلید(های) خارجی هر رابطه (در صورت وجود)

محدودیت‌های جامعیتی ناظر بر هر رابطه

طراحی با روش بالا به پایین (Top-Down)

طراحی با روش سنتز [انرمال ترسازی 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 حتی الامکان در نرمالترین فرم دست یابیم.

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



□ نمایش صحیح و واضح از خردجهان واقع باشد.



□ نمایشی صحیح از «خرد جهان واقع» به گونه ای که هیچ اشتباه معنایی در پاسخگویی به پرسش های کاربران بروز نکند.

□ تمام داده‌های کاربران قابل نمایش باشد و همه محدودیت‌های (قواعد) جامعیتی منظور شده باشد.

□ معنای هر یک از صفات از هر نوع موجودیت به درستی رعایت شده باشد.

□ کمترین میزان اختلاط اطلاعات

□ انعطاف پذیری در برابر نیازهای جدید کاربران

□ کمترین افزونگی

□ کمترین هیچمقدار

□ کمترین مشکل در عملیات ذخیره‌سازی

□ بیشترین کارایی در بازیابی

تأمین چهار ویژگی آخر به صورت همزمان، در عمل ناممکن است!



□ 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 در ادامه به آنها می‌پردازیم.

□ **فرض:** تا اطلاع ثانوی، همه صفات تک مقداری هستند و موجودیت‌ها ضعیف نیستند.



# حالت ۱: طراحی ارتباط چند به چند

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۶

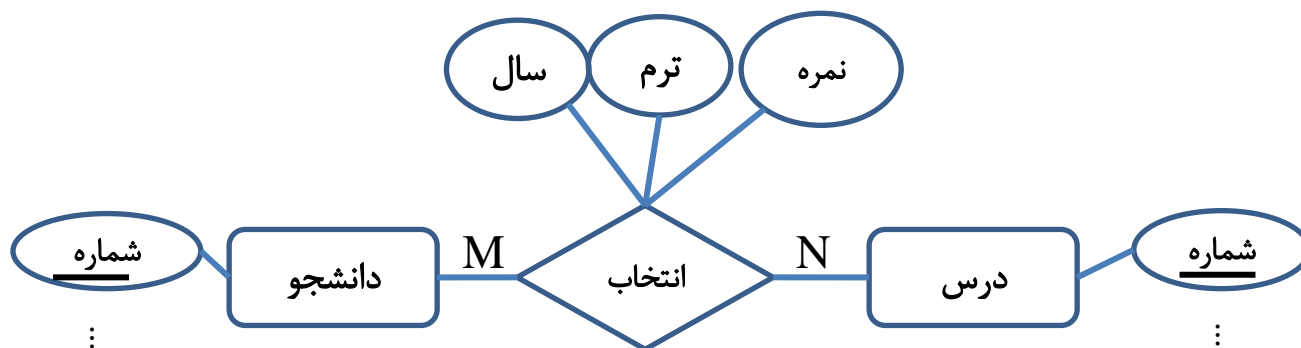
حالت ۱

درجه ارتباط:  $n=2$

چندی ارتباط:  $M:N$

سه رابطه لازم است.

طراحی در این حالت با کمتر از سه رابطه، افزونگی و هیچ‌مقداری زیادی پدید می‌آورد.



**STUD** (STID, ....)

**COR** (COID, ....)

**SCR** (STID, COID, TR, YR, GR)



## حالت ۱: طراحی ارتباط چند به چند (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۷

تعمیم حالت ۱

□ درجه:  $n > 2$

□ ابتدا فرض می‌کنیم چندی رابطه  $M:N:P:...$  است.

□  $n+1$  رابطه طراحی می‌کنیم.

□ سپس بررسی می‌کنیم که آیا محدودیت(های) خاصی روی چندی ارتباط بین بعض موجودیت‌ها وجود دارد.

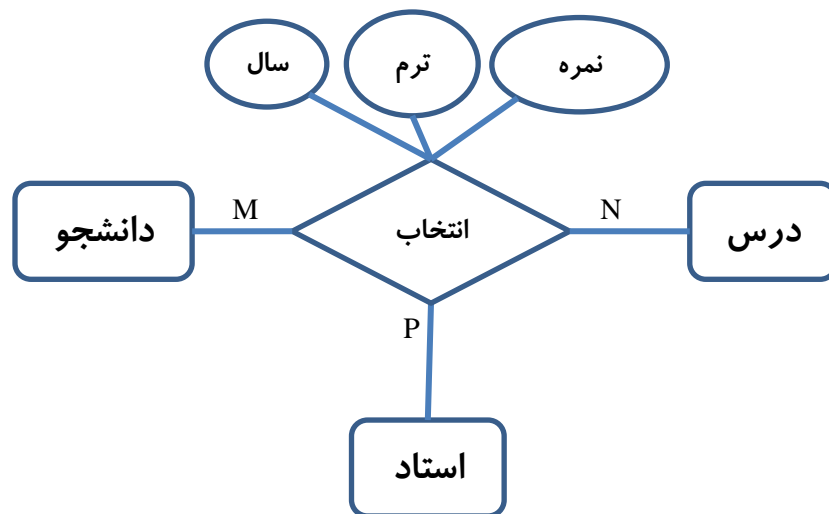
□ اگر بله، این محدودیت(ها) را در مرحله نرمالترسازی دخالت می‌دهیم. ← تعداد رابطه‌ها ممکن است بیش از  $n+1$  شود.



## حالت ۱: طراحی ارتباط چند به چند (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۸



**STUD** (STID, ....)

**COR** (COID, ....)

**PROF** (PRID, ....)

**SCP** (STID, COID, PRID, GR)

□ فرض برای «محدودیت»: یک استاد فقط یک درس را تدریس می‌کند (البته در این مورد چندی رابطه دقیق مدل نشده که این محدودیت لحاظ نشده است).

□ در این صورت باید رابطه SCP را به دو رابطه (یا بیشتر) تجزیه عمودی کنیم.

□ این محدودیت را در مرحله دوم طراحی (در مباحث آتی) دخالت می‌دهیم.





## حالت ۲: طراحی ارتباط یک به چند

بخش هفتم: طراحی پایگاه داده رابطه‌ای

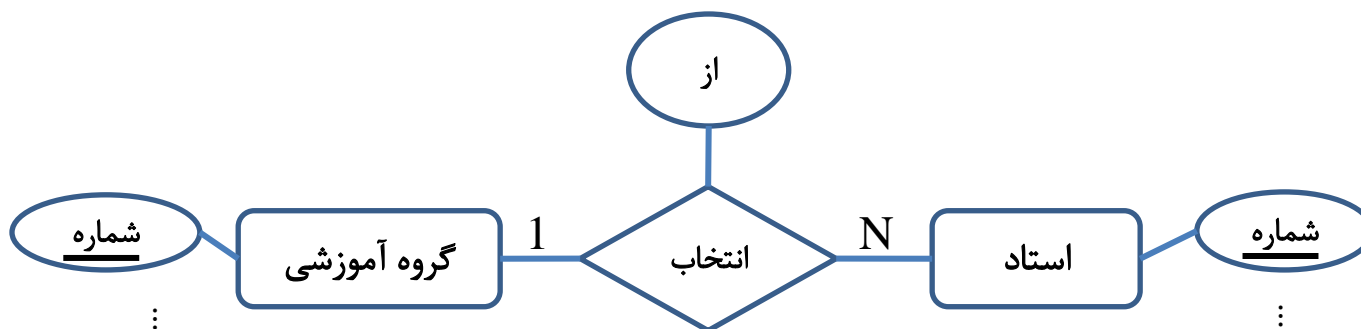
۹

حالت ۲

درجه ارتباط:  $n=2$  ☐

چندی ارتباط:  $1:N$  ☐

دو رابطه لازم است. رابطه سمت 1 به رابطه سمت N، FK می‌دهد (بیرون از کلید اصلی).



**DEPT** (DEID, DTID, ....., DPHONE)

**PROF** (PRID, PRNAME, ....., PRANK, DEID, FROM)



## حالت ۲: طراحی ارتباط یک به چند (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۱۰

□ در چه وضعی طراحی این حالت با سه رابطه قابل توجیه است؟

۱- وقتی که مشارکت سمت  $N$  در ارتباط غیرالزامی باشد (درصد مشارکت کمتر از ۳۰ درصد) و تعداد

استاد زیاد باشد، برای کاهش مقدار Null، رابطه نمایشگر ارتباط را جدا می‌کنیم.

۲- فرکانس ارجاع به خودِ ارتباط بالا باشد و به صفات دیگر با فرکانس پایین‌تری احتیاج باشد.

۳- تعداد صفات خودِ ارتباط زیاد باشد و باعث زیاد شدن درجه ارتباط PROF شود.

□ اگر مشارکت سمت  $N$  الزامی باشد، باید این محدودیت معنایی را از طریق هیچمقدارناپذیر بودن صفت کلید

خارجی (با استفاده از NOT NULL) در رابطه نمایانگر نوع موجودیت سمت  $N$ ، اعلام کرد.



## حالت ۳: طراحی ارتباط یک به یک

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۱۱

حالت ۳

درجه ارتباط:  $n=2$

چندی ارتباط: 1:1

با دو / یا سه / یا یک رابطه طراحی می‌کنیم.



در صورت طراحی با **دو** رابطه، رابطه مربوط به نوع موجودیت با مشارکت الزامی، FK می‌گیرد.

**COUR** (COID, ....., BKID)

**BOOK** (BKID, ....., BKPRICE)



## حالت ۳: طراحی ارتباط یک به یک (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۱۲

☐ وقتی با **سه** رابطه توجیه دارد :

☐ مشارکت طرفین غیرالزامی باشد

☐ تعداد شرکت کنندگان (نمونه‌ها) در ارتباط زیاد باشد

☐ درصد مشارکت در رابطه ضعیف (کمتر از ۳۰٪) باشد و نیز ملاحظات در مورد فرکانس ارجاع.

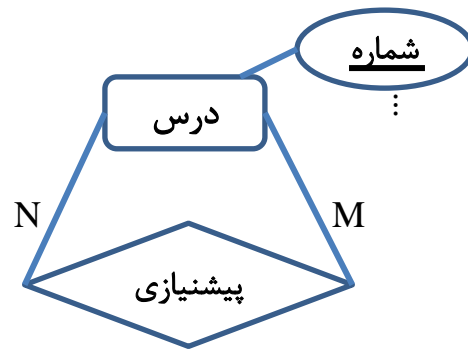
☐ وقتی با **یک** رابطه توجیه دارد :

☐ تعداد صفات موجودیت‌ها کم باشد

☐ مشارکت طرفین الزامی باشد

☐ فرکانس ارجاع به ارتباط کم باشد.

## حالت ۴



حالت خاص حالت اول ☐

درجه ارتباط:  $n=1$  ☐

چندی ارتباط:  $M:N$  ☐

دو رابطه لازم است.

**COUR** (COID, ....)

**COPRECO** (COID, PRECOID)  $\longrightarrow$  بیش از یک صفت از رابطه، از یک دامنه هستند.

**COUR**  $\longrightarrow$  **COPRECO** ☐ **گراف ارجاع:**

☐ **نتیجه:** صرف وجود ارتباط با خود، چرخه ارجاع ایجاد نمی‌شود. باید به چندی ارتباط توجه کنیم.

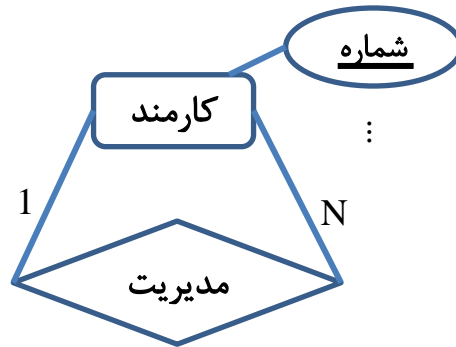


## حالت ۵: طراحی ارتباط خود ارجاع به چند

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۱۴

حالت ۵



حالت خاص حالت دوم

درجه ارتباط:  $n=1$

چندی ارتباط:  $1:N$

یک رابطه لازم است.

در این رابطه چه نکاتی وجود دارد؟

**EMPL** (EMID, ENAME, ....., EPHONE, EMGRID)



گراف ارجاع: **EMPL**

برنامه‌ای در SQL بدهید که شماره تمام مدیران در سلسله مدیریت را بدهد (با استفاده از تکنیک Recursion)

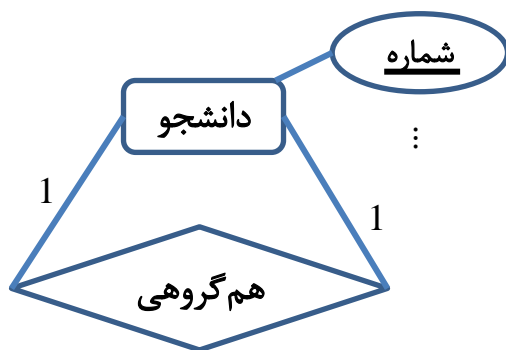


## حالت ۶: طراحی ارتباط خود ارجاع یک به یک

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۱۵

حالت ۶



□ حالت خاص حالت سوم

□ درجه ارتباط:  $n=1$

□ چندی ارتباط:  $1:1$

با یک یا دو رابطه طراحی می‌کنیم.

□ اگر مشارکت در هم‌پروژگی زیاد نباشد، از مدل II استفاده می‌کنیم.

(I) طراحی: **STPROJST** (STID, STNAME, ..., JSTID)  
P.K. C.K.

(II) طراحی: **STUD** (STID, STNAME, ..., JSTID)  
C.K. C.K.

**STJST** (STID, JSTID)  
C.K. C.K.

□ در STJST هر یک از صفات می‌توانند کلید اصلی باشند.

□ آیا طرز دیگری هم برای طراحی وجود دارد؟



## حالت ۷: طراحی موجودیت ضعیف

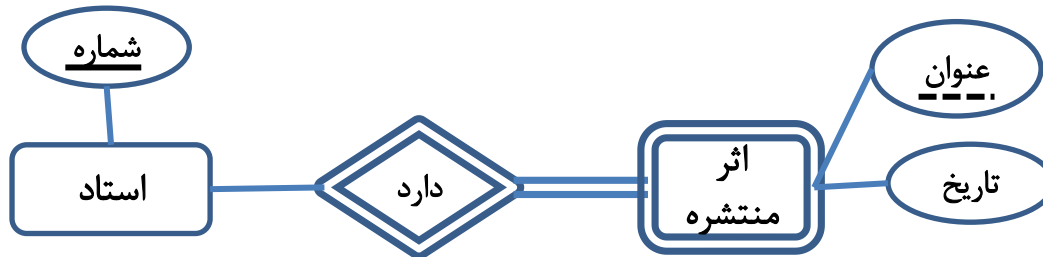
بخش هفتم: طراحی پایگاه داده رابطه‌ای

۱۶

حالت ۷

موجودیت ضعیف داریم.

دو رابطه لازم است؛ یکی برای نوع موجودیت قوی، یکی برای نوع موجودیت ضعیف و ارتباط شناسا. رابطه نمایشگر موجودیت ضعیف از موجودیت قوی FK می‌گیرد که در ترکیب با صفت ممیزه می‌شود PK.

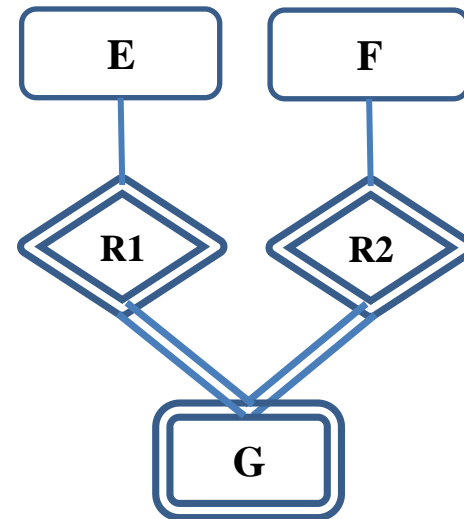
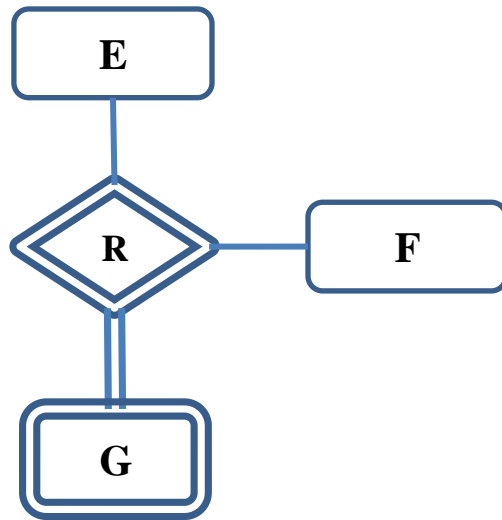


**PROF** (PRID, PRNAME, ....)

**PRPUB** (PRID, PTITLE, PTYPE, ....)



تمرین: رابطه‌های لازم برای مدل‌های داده‌ای زیر طراحی شود. ☐



در این حالت ممکن است نسبت به یک موجودیت قوی،  
یک صفت متمایز داشته باشد و نسبت به موجودیت قوی  
دیگر، صفت متمایز دیگری داشته باشد.



## حالت ۸: طراحی صفت چندمقداری

۱۸

بخش هفتم: طراحی پایگاه داده رابطه‌ای

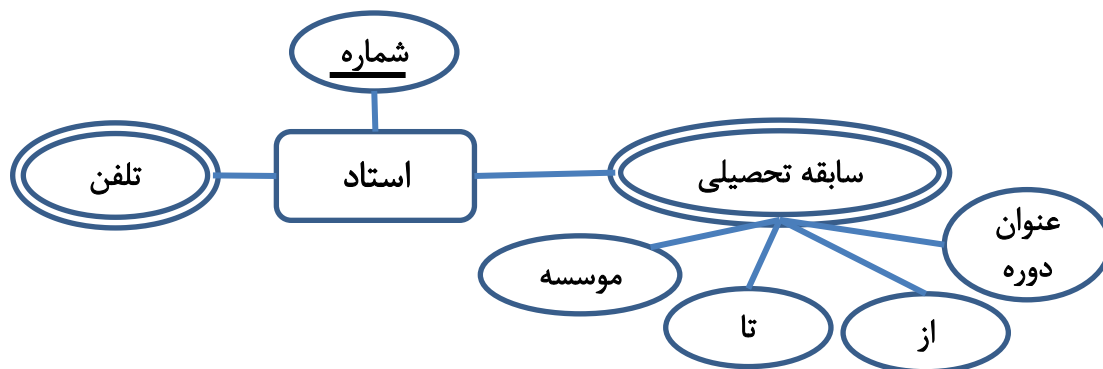
حالت ۸

□ وجود یک صفت چندمقداری برای یک نوع موجودیت.

□ سه تکنیک دارد:

۱- [تکنیک عمومی] یک رابطه برای خود نوع موجودیت و یک رابطه برای هر صفت چندمقداری.

(بنابراین اگر نوع موجودیت  $E$ ،  $m$  صفت چندمقداری داشته باشد،  $m+1$  رابطه داریم.)



**PROF** (PRID, PRNAME, ....)

**PRTEL** (PRID, PHONE)

✓ رابطه نمایشگر صفت چندمقداری از نوع

موجودیت اصلی FK می‌گیرد داخل کلید.



## حالت ۸: طراحی صفت چندمقداری (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۱۹

□ در مدلسازی، موجودیت ضعیف به صفت چندمقداری ارجحیت دارد ولی تکنیک عمومی طراحی آنها مثل هم است.

**PRHIS** (PRID, TTL, FROM, TO, INSTNAME, ....)

□ اشکال تکنیک عمومی: اگر برای نوع موجودیت اصلی اطلاعات کامل بخواهیم، باید عمل JOIN انجام دهیم که می‌تواند زمانگیر باشد.



## حالت ۸: طراحی صفت چندمقداری (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۲۰

۲- [در شرایط خاص] طراحی با یک رابطه (فرض: یک صفت چندمقداری): یک رابطه برای خود نوع موجودیت و صفت چندمقداری.

□ با فرض مشخص بودن حداکثر تعداد مقداری که صفت چندمقداری می‌گیرد، به همان تعداد صفت در رابطه در نظر می‌گیریم.

فرض: هر استاد حداکثر سه شماره تلفن دارد.



PRTELTEL (PRID, PRNAME, PRRANK, PHONE1, PHONE2, PHONE3)

□ مزیت این تکنیک: JOIN لازم ندارد و لزومی ندارد که هر استاد حتما یک شماره تلفن داشته باشد.

□ عیب این تکنیک: هیچمقدار (Null) در آن زیاد است، اگر تعداد کمی از استادان، سه شماره تلفن داشته باشند.



## حالت ۸: طراحی صفت چندمقداری (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۲۱

۳- [در شرایط خاص] طراحی با یک رابطه (یک رابطه برای خود نوع موجودیت و یک صفت چندمقداری) شامل تمام صفات نوع موجودیت و صفت چندمقداری.

دیگر صفات خود نوع موجودیت

**PRTELTEL** (PRID, PHONE, PRNAME, PRNAK, ...)

□ شرط اصلی استفاده: هر استاد حداقل یک تلفن داشته باشد.

□ شرایط دیگری که بهتر است برقرار باشد: تعداد کمی از استادها بیش از یک تلفن داشته باشند (به

دلیل افزونگی) و حتی‌الامکان تعداد صفات خود نوع موجودیت کم باشد (به دلیل افزونگی).

## حالت ۹

وجود ارتباط IS-A بین دو نوع موجودیت.

چهار تکنیک دارد:

۱- فرض: نوع موجودیت E، n زیرنوع دارد.

n+1 رابطه طراحی می‌کنیم. یک رابطه برای زیرنوع و یک رابطه برای هر یک از زیرنوع‌ها.

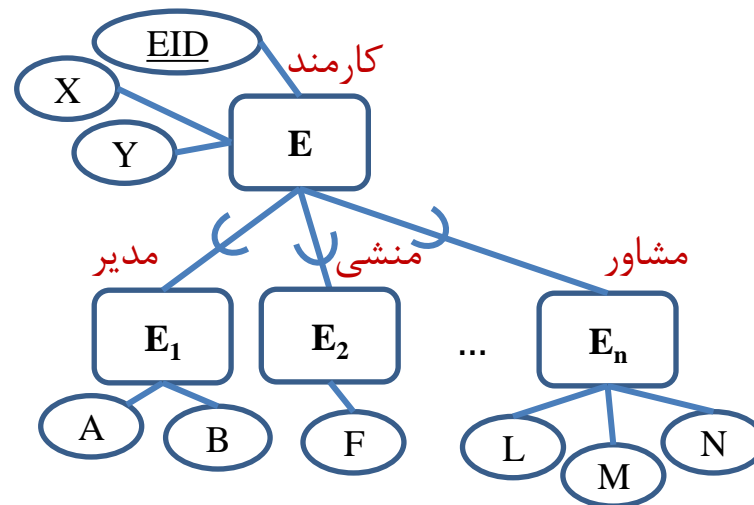
$E(\underline{EID}, X, Y)$

$E_1(\underline{EID}, A, B)$

$E_2(\underline{EID}, F)$

...

$E_n(\underline{EID}, L, M, N)$





## حالت ۹: طراحی ارتباط IS-A (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۲۳

□ مزیت این تکنیک: شرط خاصی از نظر نوع تخصیص ندارد (تکنیک‌های دیگری که مطرح می‌شود، همگی

برای شرایط خاص هستند).

□ عیب این تکنیک: اگر بخواهیم در مورد یک زیرنوع، اطلاعات کامل به دست آوریم، باید JOIN کنیم.



۲- طراحی با  $n$  رابطه: برای زیرنوع، رابطه‌ای طراحی نمی‌کنیم. بنابراین صفات مشترک باید در رابطه نمایشگر هر زیرنوع وجود داشته باشد.

$E1 (EID, X, Y, A, B)$

$E2 (EID, X, Y, F)$

...

$E_n (EID, X, Y, L, M, N)$

□ شرط لازم: باید تخصیص کامل باشد. اگر نباشد، بخشی از داده‌های محیط قابل نمایش نیستند.

□ مزیت نسبت به تکنیک اول: برای به دست آوردن اطلاعات کامل زیرنوع‌ها نیازی به JOIN نیست.

□ نکته: در این تکنیک، لزوماً افزونگی پیش نمی‌آید. اگر تخصیص هم‌پوشا باشد میزانی افزونگی پیش می‌آید.





۳- طراحی فقط با یک رابطه، با استفاده از صفت نمایشگر نوع زیرنوع‌ها

E (EID, X, Y, A, B, F, L, M, N, TYPE)

100 x1 y1 a1 b1 ? ? ? ? مدیر

200 x2 y2 ? ? ? l2 m2 n2 مشاور

□ شرط استفاده از این تکنیک: تخصیص مجزا باشد؛ یعنی یک نمونه کارمند، جزء نمونه‌های حداکثر یک

زیرنوع باشد.

□ مزیت این تکنیک: برای به دست آوردن اطلاعات کامل زیرنوع‌ها نیازی به JOIN نیست.

□ عیب این تکنیک: هیچ مقدار (Null) زیاد دارد و درجه رابطه زیاد است.

**۴- طراحی فقط با یک رابطه، با استفاده از آرایه بیتی؛ هر بیت نمایشگر نوع یک زیرنوع. در واقع برای**

نمایش هر نمونه موجودیت، بسته به اینکه در مجموعه نمونه‌های کدام زیرنوع باشد، بیت مربوطه‌اش را ۱ می‌کنیم.

**E** (EID, X, Y, A, B, F, L, M, N,  $\overbrace{\text{TB1, TB2, ..., TBn}}^{\text{آرایه بیت‌ی}}$ )

100	x1	y1	1	0	0
200	x2	y2	0	1	0

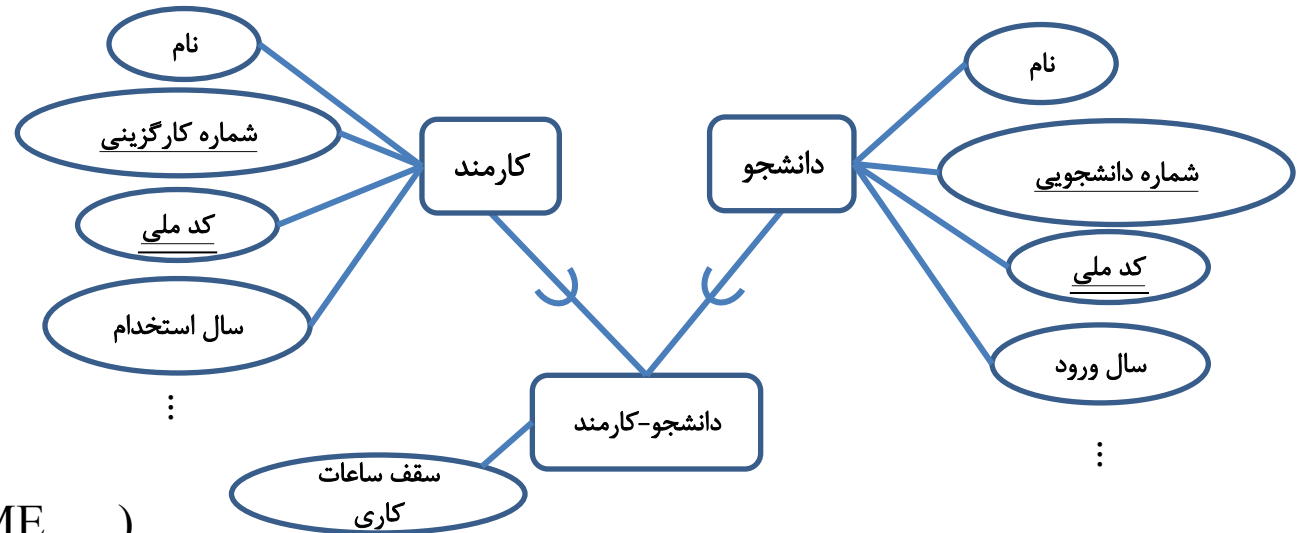
شرط استفاده از این تکنیک: وقتی تخصیص هم‌پوشا باشد (سایر شرایط همانها که در تکنیک ۳ گفته

شد).

## حالت ۱۰

وجود ارث‌بری چندگانه بین یک زیرنوع با چندزیرنوع

اگر زیرنوع،  $n$  زیرنوع داشته باشد، رابطه نمایشگر زیر حداقل  $n$  کلید کاندید دارد.



**STUD** (STID, STNAME, ...)

**EMPL** (EID, ENAME, ...)

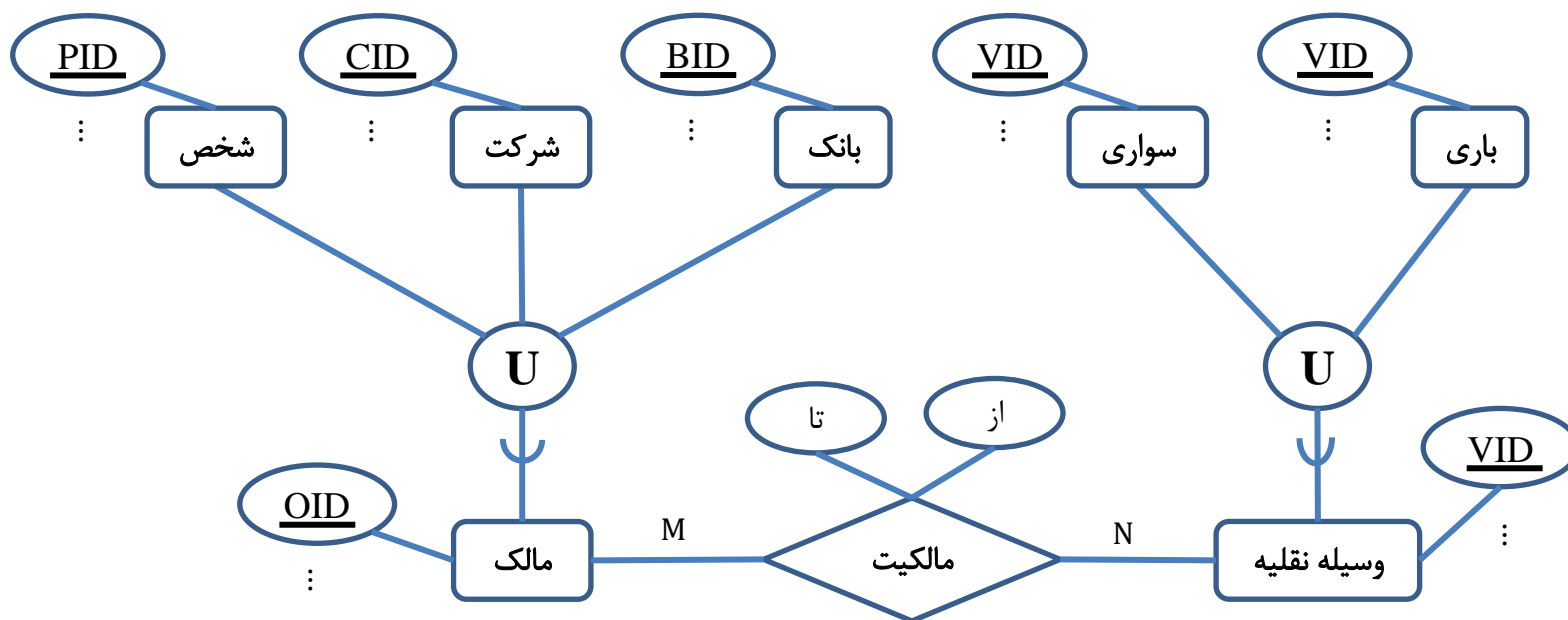
**STEM** (STID, EID, MAXW)

آیا ممکن است برای زیرنوع اصلاً رابطه طراحی نکنیم؟



□ نوع موجودیت E، زیرنوع U-Type (دسته یا Category) n زیرنوع است.

n+1 رابطه طراحی می‌کنیم.





## حالت ۱۱: طراحی زیرنوع اجتماع (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۲۹

□  $n+1$  رابطه

□ اگر شناسه زبرنوع‌ها از دامنه‌های متفاوت باشد، رابطه نمایشگر زیرنوع، FK می‌دهد به رابطه‌های نمایشگر زبرنوع‌ها، خارج از کلید.

□ اگر شناسه زبرنوع‌ها از یک دامنه باشد، کلید رابطه نمایشگر زیرنوع، همان کلید رابطه‌های نمایشگر زبرنوع‌ها است و مفهوم FK به طور صریح مطرح نیست.

**COMP** (CID, ....., OID) (برای IS-A هم همین نکته مطرح است.)

**BANK** (BID, ....., OID)

**OWNER** (OID, .....) ← چون دامنه کلیدهای زبرنوع‌ها یکسان نیست، خودمان کلید ساختگی می‌گذاریم.

**VEHIC** (VID, ....)

**OWNS** (OID, VID, F, T, ....)

**SAVARY** (VID, N, ....)

**BARY** (VID, T, ....)

آیا طرز طراحی دیگری وجود دارد؟

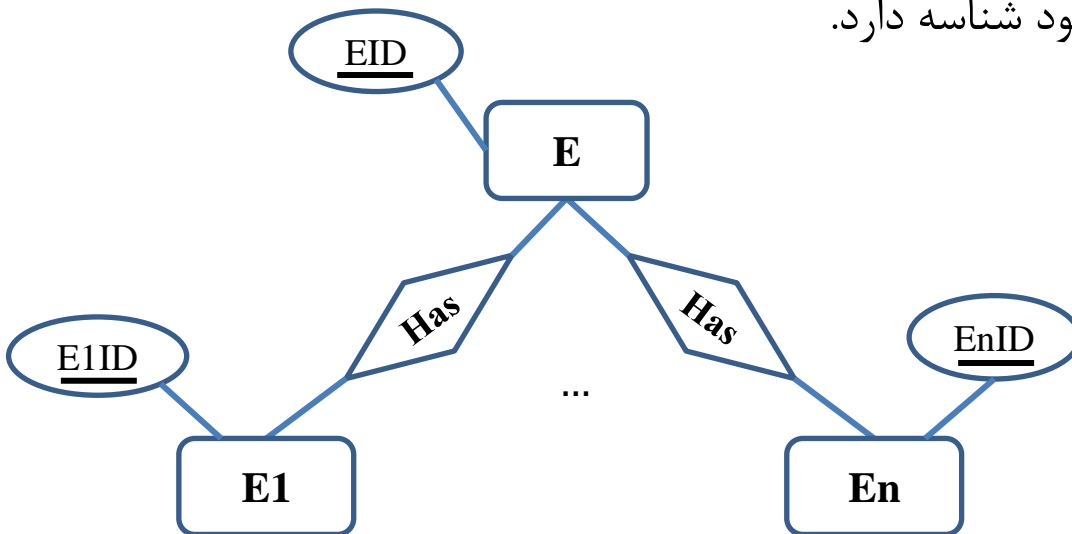


## حالت ۱۲

وجود ارتباط IS-A-PART-OF

اگر نوع موجودیت کل،  $n$  نوع موجودیت جزء داشته باشد، تعداد  $n+1$  رابطه طراحی می‌کنیم.

توجه داریم که نوع موجودیت جزء از خود شناسه دارد.



$E(\underline{EID}, \dots)$

$E1(\underline{E1ID}, \underline{EID}, \dots)$

....

$En(\underline{EnID}, \underline{EID}, \dots)$

آیا طرز طراحی دیگری وجود دارد؟ در چه شرایطی؟





# حالت ۱۳: طراحی تکنیک Aggregation

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۳۱

حالت ۱۳

استفاده از تکنیک Aggregation در مدلسازی

ابتدا نوع موجودیت انتزاعی (بخش درون مستطیل خط‌چین) را طراحی می‌کنیم (با توجه به درجه و چندی ارتباط). سپس بخش بیرون آن را (باز هم با توجه به چندی ارتباط و درجه آن).

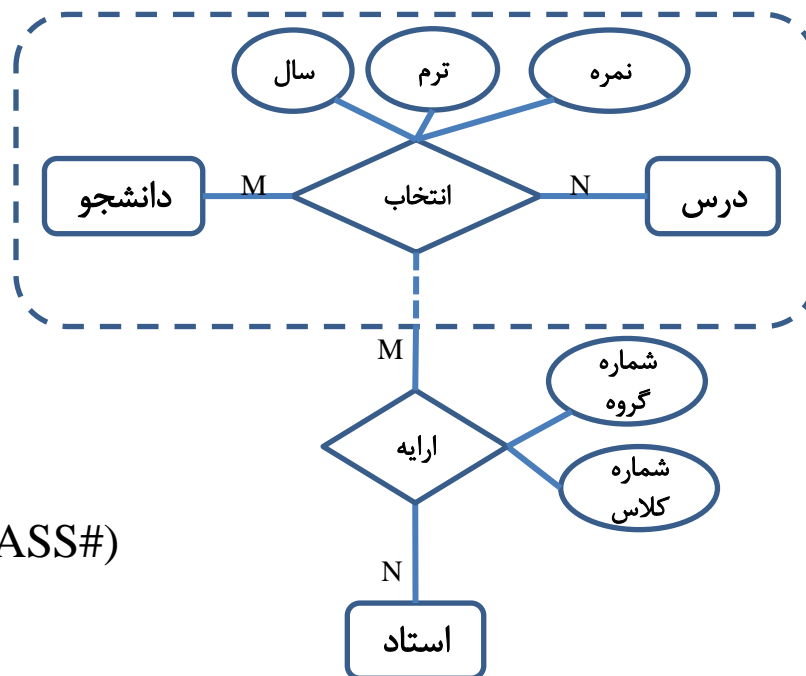
**STUD** (STID, ....)

**COUR** (COID, ....)

**SCR** (STID, COID, TR, YR, GR)

**PROF** (PRID, ....)

**OFFERING** (STID, COID, PRID, GR#, CLASS#)





## حالت ۱۳: طراحی تکنیک Aggregation (ادامه)

بخش هفتم: طراحی پایگاه داده رابطه‌ای

۳۲

این تکنیک چگونه کارایی سیستم را افزایش می‌دهد (نسبت به طراحی با یک ارتباط سه-تایی)؟ ☐

اگر مراجعه به ارتباط «انتخاب» بالا باشد و فرکانس ارجاع به ارتباط «ارائه» پایین باشد، سیستم با این ☐

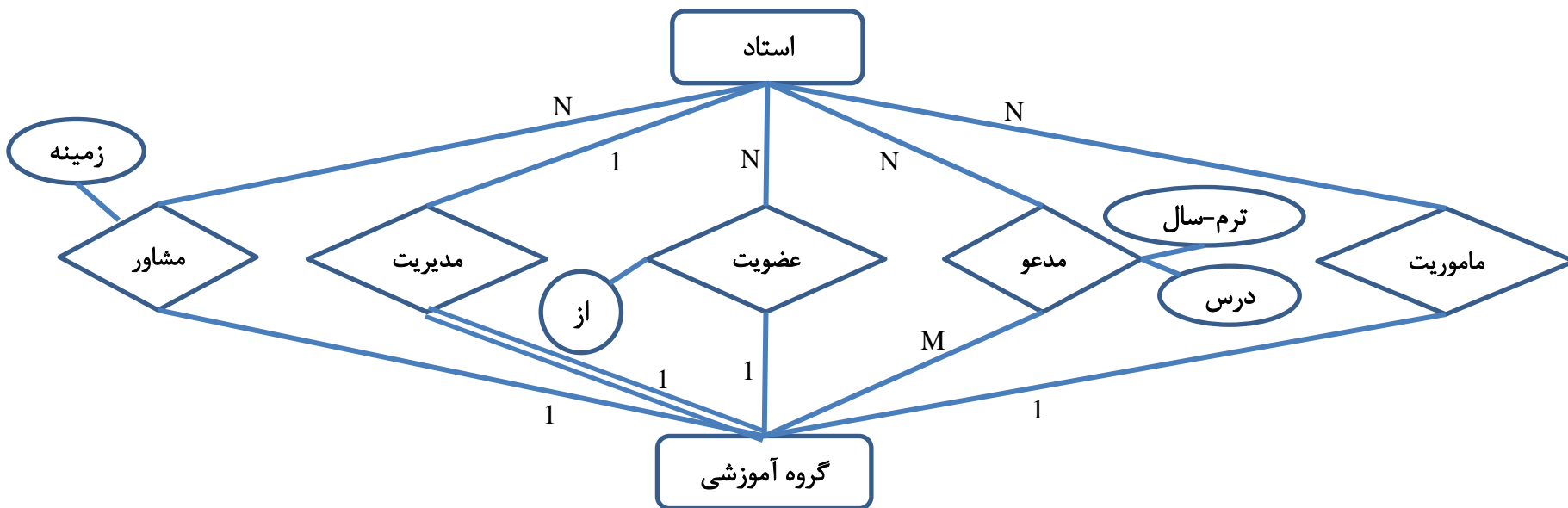
طراحی کاراتر عمل می‌کند.



## حالت ۱۴

در صورتیکه چند ارتباط مثلاً بین دو نوع موجودیت برقرار باشد.

هر ارتباط را با توجه به وضع آن از نظر درجه و چندی ارتباط طراحی می‌کنیم. اما برای کاهش احتمال اشتباه در طراحی توصیه می‌شود اول ارتباطهای  $M:N$ ، سپس  $1:N$  و در آخر  $1:1$  را طراحی می‌کنیم.





## حالت ۱۴: طراحی با وجود چند ارتباط (ادامه)

۳۴

بخش هفتم: طراحی پایگاه داده رابطه‌ای

**DEPT** (DEID, ....., DPHONE, PRID)

**PROF** (PRID, ....., PRRANK, MDEID, SUB, MEMDEID, FROM, CDEID, INT)

زمینه مشاور از عضویت موضوع ماموریت

**INVITED** (DEID, PRID, YR, TR)

سه کلید خارجی از یک دامنه

همین سیستم حداکثر با هفت رابطه نیز قابل طراحی است.



## حالت ۱۵ : ارتباط سلسله مراتبی

۳۵

بخش هفتم: طراحی پایگاه داده رابطه‌ای

**تمرین:** فرض می‌کنیم بین  $n$  نوع موجودیت، هر یک نمایش داده شده با یک نوع رکورد، ارتباط

سلسله‌مراتبی وجود داشته باشد (بر اساس ساختار سلسله‌مراتبی HDS). مطلوب است طراحی این محیط در مدل رابطه‌ای.

**E1 REC**

E1ID ....

**E2 REC**

E2ID ....

**En REC**

EnID ....



□ ایده اصلی: 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.

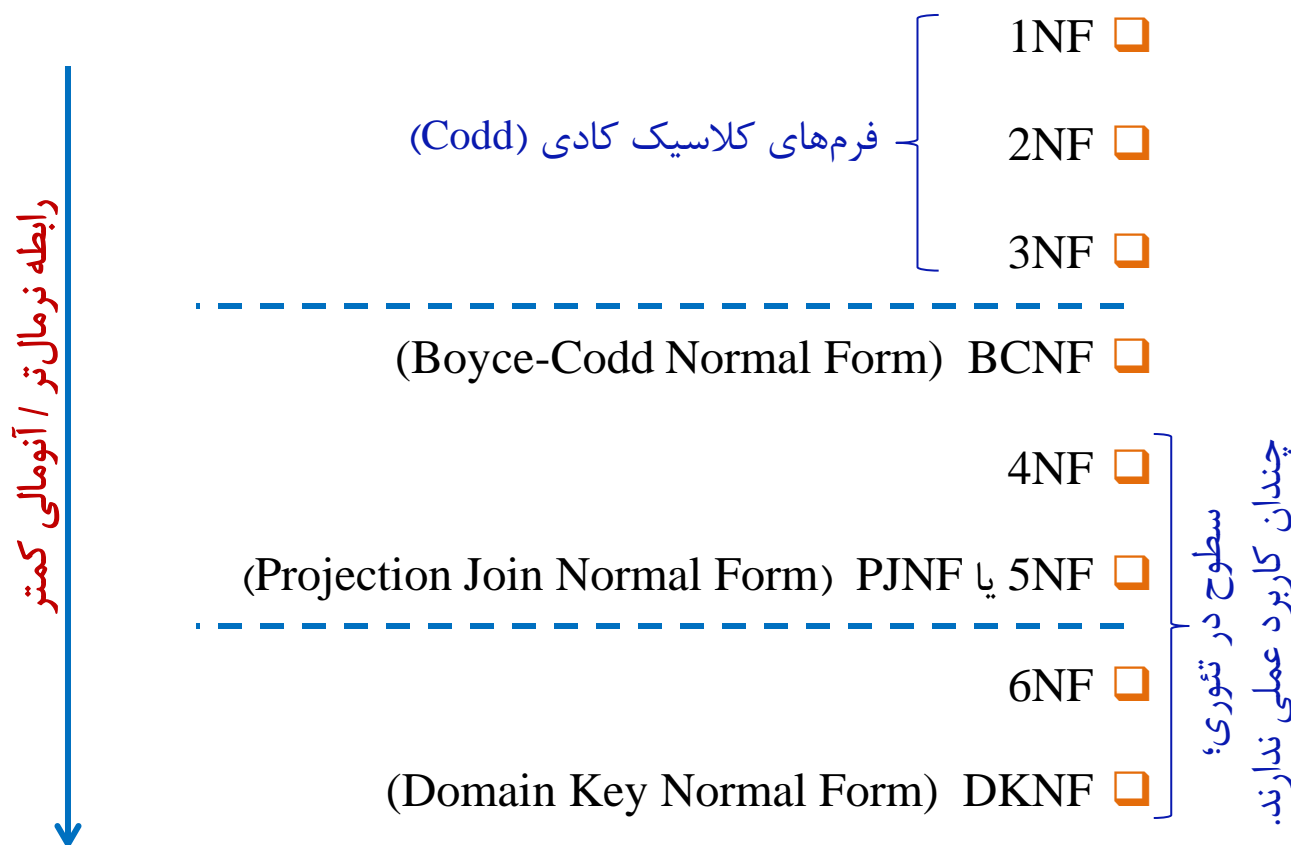
□ **آنومالی در بهنگام‌سازی:** بروز فزون کاری.

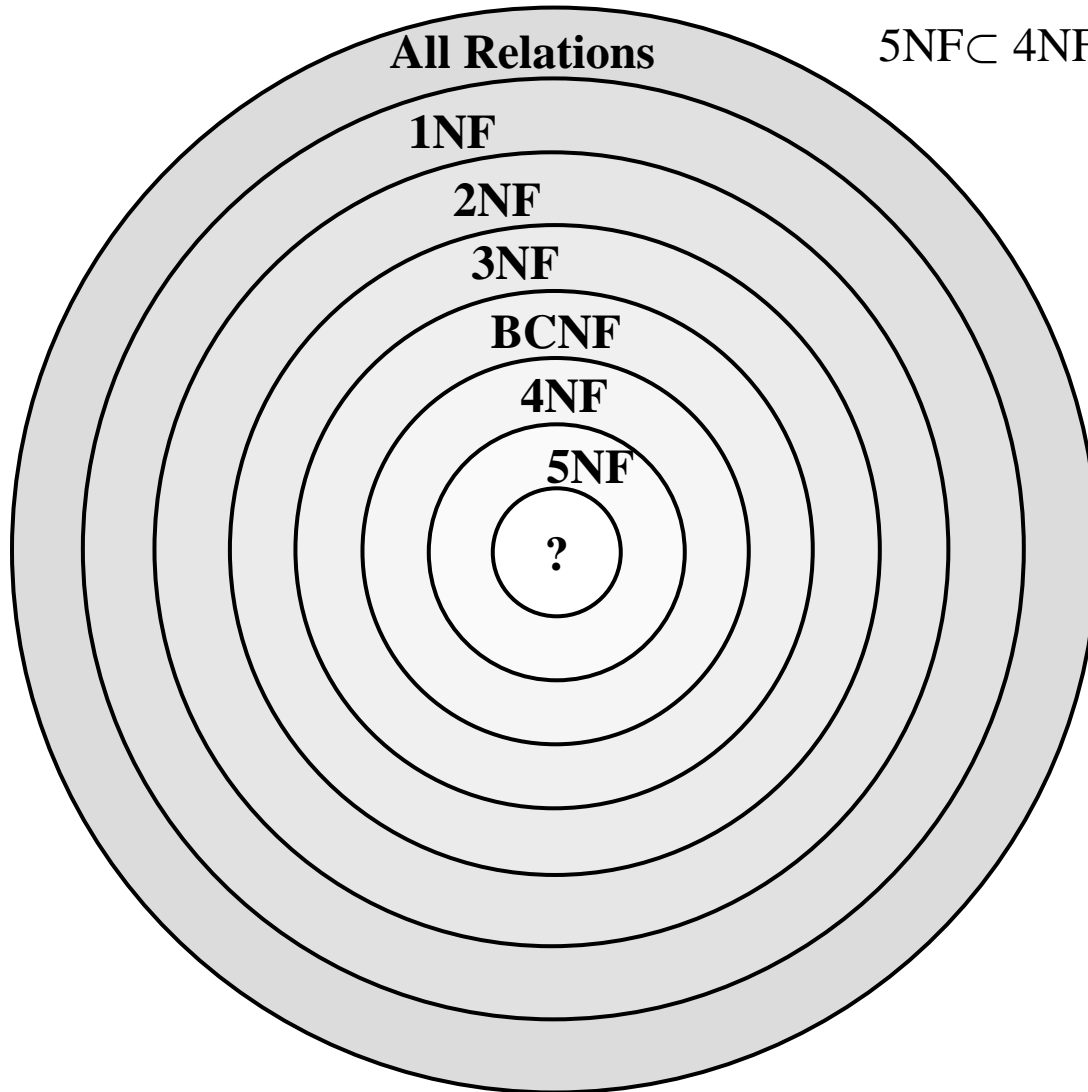
□ پس باید رابطه را نرمال‌تر کرد.



□ 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 رابطه‌ای که BCNF باشد، ☐

مثلا 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:** صفت R.B با 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  $t_i$  و  $t_j$  دو تاپل از R باشند، در این صورت:

$$A \rightarrow B \Leftrightarrow \forall t_i, t_j \in R, \quad \text{IF } t_i.A = t_j.A \text{ THEN } t_i.B = t_j.B$$

:  $A \rightarrow B$

A Determines B ■

■ A ، B را تعیین می کند.

آیا داریم:



**R( A , B , C )**

$a_1$	$b_1$	$c_1$
$a_1$	$b_1$	$c_2$
$a_2$	$b_2$	$c_2$
$a_3$	$b_3$	$c_3$
$a_4$	$b_2$	$c_3$

$a_1 \rightarrow b_1$

$a_1 \rightarrow b_1$

$a_2 \rightarrow b_2$

$a_3 \rightarrow b_3$

$a_4 \rightarrow b_4$

$a_1 \rightarrow c_1$

$a_1 \rightarrow c_2$

...

بله  $A \rightarrow B$  ؟

خیر  $A \rightarrow C$  ؟

خیر  $B \rightarrow A$  ؟

خیر  $B \rightarrow C$  ؟





نکات: 

(۱) صفات طرفین FD می‌توانند 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:  $K \rightarrow G$ .

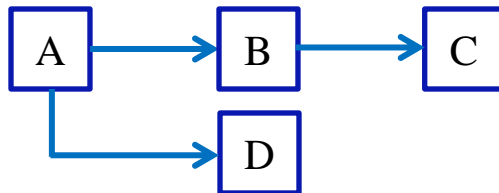


(۵) نمایش FDهای رابطه R

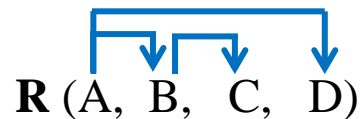
- به صورت یک مجموعه:

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

- 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).

STID → STJ: یک دانشجو فقط در یک رشته تحصیل می‌کند.



STJ → STD: یک رشته فقط در یک دانشکده ارائه می‌شود.

STID → STD: یک دانشجو فقط در یک دانشکده تحصیل می‌کند.

□ **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:



## قواعد استنتاج آرمسترانگ ☐

- 1- if  $B \subseteq A$  then  $A \rightarrow B \Rightarrow A \rightarrow A$  (قاعده انعکاس)
- 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 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:** سه قاعده اول به آسانی قابل اثبات هستند و قواعد دیگر از روی همانها اثبات می‌شوند.

□ **قاعده یگانگی عمومی (General Unification):**

$if A \rightarrow B \text{ AND } C \rightarrow D \text{ THEN } A \cup (C - B) \rightarrow (B, D)$  □

چرا به این قاعده، قاعده یگانگی عمومی گویند؟

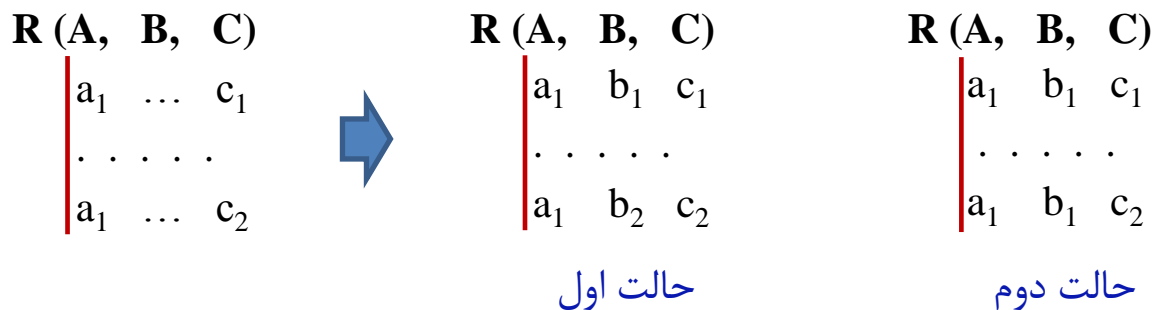




تمرین: قاعده ۲ را اثبات کنید (با استفاده از برهان خلف).

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 می‌شود. 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 کاهش‌ناپذیر FD های یک رابطه

سه شرط دارد:

۱- هیچ FD در آن افزونه نباشد.

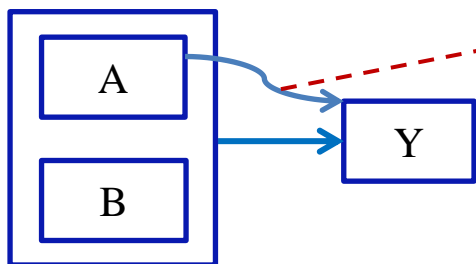
۲- سمت راست هر FD، صفت ساده باشد.

۳- سمت چپ هر FD، خود کاهش‌ناپذیر باشد: در وابستگی تابعی  $X \rightarrow Y$ ،  $X$  را کاهش‌ناپذیر (و

وابستگی  $X \rightarrow Y$  را کامل) گوییم، هرگاه  $Y$  با هیچ زیرمجموعه از  $X$  (غیر از خود  $X$ )، FD نداشته باشد.

در غیر اینصورت  $X$  را کاهش‌پذیر گوییم و وابستگی  $X \rightarrow Y$  را ناکامل گوییم.

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



$$\begin{cases} (A, B) \rightarrow Y \\ A \rightarrow Y \end{cases} \Rightarrow \text{FD ناکامل}$$



□ **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 می‌گیریم  $(A,B) \rightarrow (Y,B)$

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

کجای؟ مجموعه کاهش‌ناپذیر چه کاربردی دارد؟

تاریخ وابستگی تابعی با واسطه (TFD): اگر  $A \rightarrow B$ ،  $B \rightarrow C$  و  $A \not\rightarrow B$ ، می‌گوییم 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 صرفاً 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 نباشد و جزء PK 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)**

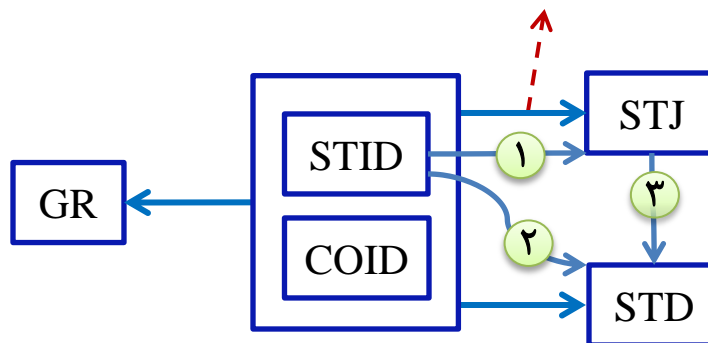
777	CO1	Phys	D11	19
777	CO2	Phys	D11	16
777	CO3	Phys	D11	11
888	CO1	Math	D12	16
888	CO2	Math	D12	18
444	CO1	Math	D12	13
555	CO1	Phys	D11	14
555	CO2	Phys	D11	12

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 نرمال تر شود.

☐ آنومالی‌های رابطه **R**:

۱- در درج:

درج کن این فقره اطلاع درمورد یک دانشجو را:  $\langle '666', 'chem', 'D16' \rangle$

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:  $\langle '444', 'CO1', 13 \rangle$

حذف انجام می‌شود اما اطلاع ناخواسته 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)

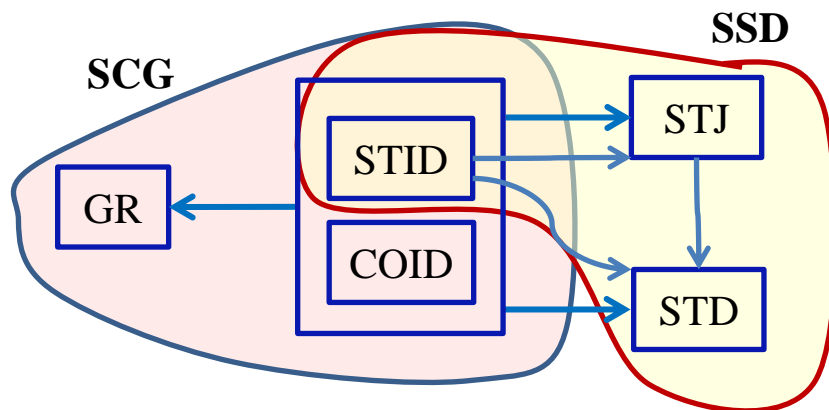
777	CO1	19
777	CO2	16
777	CO3	11
888	CO1	16
888	CO2	18
444	CO1	13
555	CO1	14
555	CO2	12

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

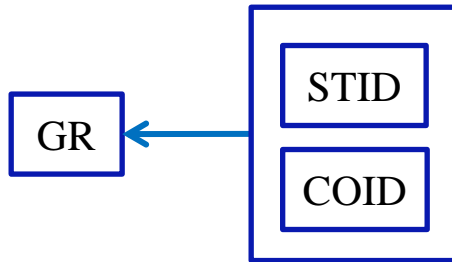


**SSD** (STID, STJ, STD)

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



SCG



رابطه‌های جدید آنومالی‌های R را ندارند: □

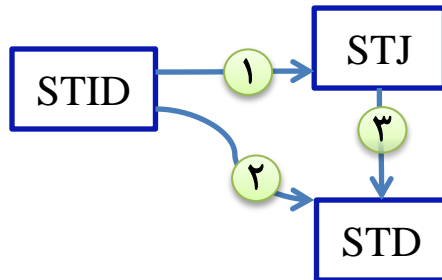
۱- درج کن:  $\langle '666', 'chem', 'D16' \rangle$

بدون مشکل در SSD درج می‌شود.

۲- حذف کن:  $\langle '444', 'CO1', 13 \rangle$

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. (پیوند از نوع CK-FK)

$$R = SCG \bowtie SSD$$



□ در حالت کلی اگر  $R_1, R_2, \dots, R_n$  پرتوهای دلخواه از  $R$  باشند، داریم (ممکن است تاپل‌های افزونه بروز کند):

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

□ 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) گوییم.

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



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

□ **آنومالی‌های SSD:** SSD (STID, STJ, STD)

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>'

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.

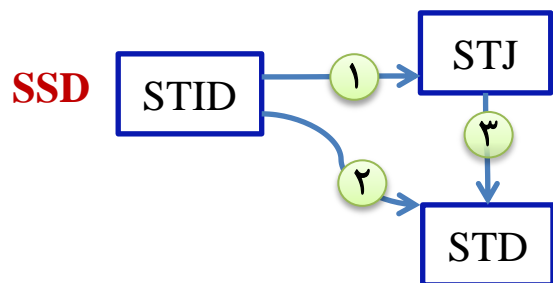
SSD باید نرمال‌تر شود.





## □ 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	Phys	D11
888	Math	Math	D12
444	Math		
555	Phys		

□ افزودن کمی کم شد!

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

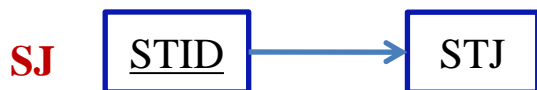


## فرم‌های نرمال کلاسیک کادی (ادامه)

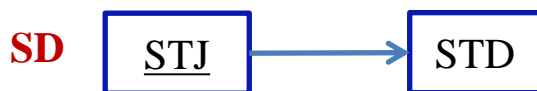
۶۱

بخش هفتم: طراحی پایگاه داده رابطه‌ای

☐ این رابطه‌ها در 3NF هستند.



☐ اولاً در 2NF هستند.



☐ ثانیاً FD با 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».



## □ تجزیه خوب (Nonloss/Lossless Decomposition)

۱- بی‌حشو: 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- بی‌حذف: در پیوند پرتوها هیچ تاپلی حذف نشود (صفت یا صفات پیوند هیچمقدار نباشند).

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

□ در بیشتر متون کلاسیک، بحث تجزیه خوب، تحت عنوان تجزیه بی‌کاست یا بی‌گمشدگی

(Nonloss/Lossless Decomposition) مطرح شده است، که چندان مناسب به نظر نمی‌رسد، مگر آنکه فرض کنیم که منظور همان بی‌حشو و حافظ وابستگی‌های تابعی بودن است (و دو ویژگی دیگر تجزیه خوب را پیش فرض تجزیه خوب بدانیم).

## □ قضیه ريسانن (Rissanen):

□ رابطه  $R$  به دو پرتوش  $(R_1$  و  $R_2)$  تجزیه خوب می‌شود، 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  $\Leftarrow$  بی‌حشو بودن

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

استنتاج شوند  $\Leftarrow$  حافظ  $FD$ ها

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

باشد، در اینصورت تجزیه خوب چنین است:  $R_1(A, B)$  و  $R_2(B, C)$ .

□ در اینجا  $B$  در رابطه دوم کلید کاندید است، چون همه صفات به آن وابستگی تابعی دارند و کاهش‌پذیر

هم نیست.



□ مثال: رابطه SSD را در نظر می‌گیریم. این رابطه به سه شکل به پرتوهای دوگانی تجزیه می‌شود.

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 کاهش‌ناپذیر، CK باشد.

☐ در 3NF، تنها باید دترمینان رابطه PK باشد.

☐ چون رابطه می‌تواند بیش از یک CK داشته باشد، BCNF از 3NF قوی‌تر است.



رابطه‌های زیر در BCNF هستند.



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

SJ STID → STJ

SD STJ → STD

SCG [ STID  
COID ] → GR

C.K.



☐ BCNF از 3NF قوی‌تر است.  $\Leftarrow$  رابطه می‌تواند 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.

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

☐ **I-II 2:** CKها هم‌پوشا باشند.  $\Leftarrow$  اگر R در 3NF باشد، لزوماً در BCNF نیست.

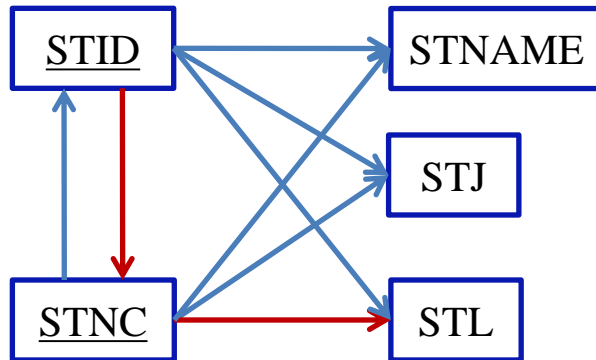


برای حالت II-1



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

دو دترمینان، هر دو هم CK هستند.

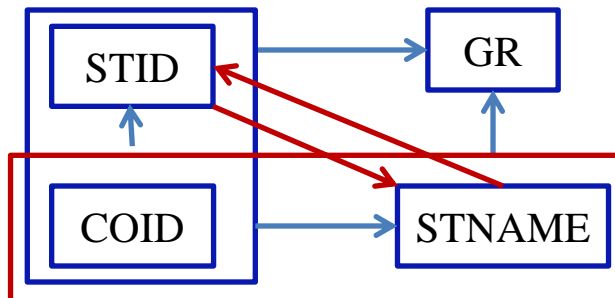


SCNG (STID, COID, STNAME, GR)  
C.K. C.K.

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؟

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

□ 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 نمی‌کند 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 می‌کنیم طراح رابطه زیر را طراحی کرده است.

SCNG (ST#, CO#, PR#)  
C.K. C.K.

□ 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 رابطه‌های حاصل

BCNF باشد.

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

**قضیه هیت (Heath):** در رابطه  $R(A, B, C)$  که در آن  $A$ ،  $B$  و  $C$  سه مجموعه از صفات هستند، اگر  $A \rightarrow B$

آنگاه تجزیه  $R$  به دو پرتو  $R_1(A, B)$  و  $R_2(A, C)$  تجزیه بی‌کاست (Nonloss) است.



## تعریف زانیلو از BCNF: □

□ رابطه‌ی  $R$  با مجموعه FD هایش  $(F)$  در BCNF است اگر و فقط اگر به ازای هر FD در  $F^+$ ، مانند  $X \rightarrow A$ ، یکی از دو شرط زیر برقرار باشد:

اگر به جای  $F^+$  بگذاریم  $F$ ، چه تغییری در شرایط پیش می‌آید؟



1. آن FD نامهم باشد.

2.  $X$  سوپرکلید باشد.

## تعریف زانیلو از 3NF: □

□ رابطه‌ی  $R$  با مجموعه FD هایش  $(F)$  در 3NF است اگر و فقط اگر به ازای هر FD در  $F^+$ ، مانند  $X \rightarrow A$ ، یکی از سه شرط زیر برقرار باشد:

1. آن FD نامهم باشد.

2.  $X$  سوپرکلید باشد.

3.  $A$  جزء کلید باشد.





## وابستگی چندمقداری (MVD):



□ در رابطه  $R(A, B, C)$  (رابطه با سه صفت یا سه مجموعه صفات)، صفت  $B$  با صفت  $A$ ، MVD دارد

اگر و فقط اگر به ازای یک مقدار  $A$ ، مجموعه‌ای به خوبی تعریف شده (خوش تعریف) از مقادیر  $B$  متناظر باشد که به تغییرات  $C$  وابسته نباشد.

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

□ طرز نمایش:  $A \twoheadrightarrow B$




وابستگی چندمقداری بین A و B :



$R(A, B, C)$

$a_1$	$\begin{Bmatrix} b_1 \\ b_2 \\ b_3 \end{Bmatrix}$	$c_1$
$a_1$	$\begin{Bmatrix} b_1 \\ b_2 \\ b_3 \end{Bmatrix}$	$c_2$
$a_2$	$\begin{Bmatrix} b_1 \\ b_7 \end{Bmatrix}$	$c_3$

$\rightarrow A \twoheadrightarrow B$

نکته : پدیده MVD بیان فرمال صفت چندمقداری است. 



4NF:

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



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# )

$PR_1$	$\begin{Bmatrix} 777 \\ 888 \\ 444 \end{Bmatrix}$	$\begin{Bmatrix} R_1 \\ R_2 \end{Bmatrix}$
$PR_2$	$\begin{Bmatrix} 777 \\ 666 \end{Bmatrix}$	$\{R_3\}$
...	...	...



NNCTB( C# , T# , B# )

$c_1$	$\begin{Bmatrix} t_1 \\ t_2 \\ t_3 \end{Bmatrix}$	$\begin{Bmatrix} b_1 \\ b_2 \end{Bmatrix}$
$c_2$	$\begin{Bmatrix} t_4 \\ t_7 \end{Bmatrix}$	$\begin{Bmatrix} b_3 \\ b_4 \\ b_5 \end{Bmatrix}$
...	...	...

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





CTB( C# , T# , B# )

$c_1$	$t_1$	$b_1$
$c_1$	$t_1$	$b_2$
$c_1$	$t_1$	$b_3$
$c_1$	$t_2$	$b_1$
$c_1$	$t_2$	$b_2$
$c_1$	$t_2$	$b_3$
$c_1$	$t_3$	$b_1$
$c_1$	$t_3$	$b_2$
$c_1$	$t_3$	$b_3$
$c_2$	$t_4$	$b_3$
...	...	...

مثال  
فرم نرمال شده ی مثال پیش :

☐ 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 بهنگام‌سازی:** شماره  $c_1$  را عوض کن.

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

نمی‌توانیم بگوییم چون کلید نداریم نمی‌توانیم 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 مجموعه‌ای

رابطه CTB باید تجزیه شود تا رابطه‌های 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 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 <sub>1</sub>	t <sub>1</sub>
c <sub>1</sub>	t <sub>2</sub>
c <sub>1</sub>	t <sub>3</sub>
c <sub>2</sub>	t <sub>4</sub>
c <sub>2</sub>	t <sub>2</sub>

CB (C#, B#)

c <sub>1</sub>	b <sub>1</sub>
c <sub>1</sub>	b <sub>2</sub>
c <sub>2</sub>	b <sub>3</sub>
c <sub>2</sub>	b <sub>5</sub>
c <sub>2</sub>	b <sub>7</sub>
c <sub>1</sub>	b <sub>8</sub>

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 بی‌کاست

(Nonloss) می‌شود اگر و فقط اگر «  $A \rightarrow\rightarrow B$  ».

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

❑ آیا می‌توان گفت مفهوم MVD تعمیم مفهوم FD است؟ به بیان دیگر، آیا FD حالت خاصی از MVD است؟

❑ FD حالت خاصی از MVD است که در آن مجموعه‌های خوش‌تعریف، تک‌عنصری هستند.

❑ همچنین این استنتاج منطقی را هم داریم:

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



### تعریف زانیلو از 4NF:

 رابطه‌ی  $R$  با مجموعه MVD هایش  $(F)$  در 4NF است اگر و فقط اگر به ازای هر MVD در  $F^+$ ,

مانند  $A \twoheadrightarrow X$ ، یکی از دو شرط زیر برقرار باشد:

1. آن MVD نامهم باشد.

2.  $X$  سوپرکلید باشد.



❑ **نکته:** بحث 4NF از یک دیدگاهی اصلاً می‌تواند موضوعیت نداشته باشد. زیرا رابطه‌ای که BCNF باشد و

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

ابتدا می‌توان رابطه‌های جداگانه طراحی کرد.

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

MVD ادغام شده؟





**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 این  $n$  پرتو باشد (و نه کمتر).

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



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

□ 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.

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

$$\left\{ \begin{array}{l} \text{STN} (\text{STID}, \text{STNAME}) \\ \text{SJ} (\text{STID}, \text{STJ}) \\ \text{SL} (\text{STID}, \text{STL}) \end{array} \right\} \Rightarrow \text{STUD} = [\text{JD}] * (\text{STN}, \text{SJ}, \text{SL})$$
 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، که هیچکدام سوپرکلید نیستند.

**PCD (PR#, CO#, D#)**

PR <sub>1</sub>	co <sub>1</sub>	d <sub>2</sub>
PR <sub>1</sub>	co <sub>2</sub>	d <sub>1</sub>
PR <sub>2</sub>	co <sub>1</sub>	d <sub>1</sub>
PR <sub>1</sub>	co <sub>1</sub>	d <sub>1</sub>

□ استاد PR# درس CO# را در دانشکده D# ارائه می‌دهد.

رابطه SPJ تمام کلید است.  $\Leftarrow$  حداقل BCNF

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

S <sub>1</sub>	P <sub>1</sub>	J <sub>1</sub>
S <sub>1</sub>	P <sub>1</sub>	J <sub>2</sub>
S <sub>1</sub>	P <sub>2</sub>	J <sub>1</sub>
S <sub>2</sub>	P <sub>1</sub>	J <sub>1</sub>

MVD ندارد.  $\Leftarrow$  4NF است.



بخش هفتم: طراحی پایگاه داده رابطه‌ای

فرض می‌کنیم بخش هفتم این رابطه را تجزیه کنیم:

SP (S#, P#)

S <sub>1</sub>	P <sub>1</sub>
S <sub>1</sub>	P <sub>2</sub>
S <sub>2</sub>	P <sub>1</sub>

PJ (P#, J#)

P <sub>1</sub>	J <sub>1</sub>
P <sub>1</sub>	J <sub>2</sub>
P <sub>2</sub>	J <sub>1</sub>



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

S <sub>1</sub>	P <sub>1</sub>	J <sub>1</sub>
S <sub>1</sub>	P <sub>1</sub>	J <sub>2</sub>
S <sub>1</sub>	P <sub>2</sub>	J <sub>1</sub>
S <sub>2</sub>	P <sub>1</sub>	J <sub>2</sub>
S <sub>2</sub>	P <sub>1</sub>	J <sub>1</sub>

تاپل حشو

SJ (S#, J#)

S <sub>1</sub>	J <sub>2</sub>
S <sub>1</sub>	J <sub>1</sub>
S <sub>2</sub>	J <sub>1</sub>

این رابطه JD به دو پرتوش ندارد.

یک پرتو دیگر هم می‌گیریم:

$$SPJ \neq [JD]^*(SP, PJ)$$



SPJ (S#, P#, J#)

۱	S <sub>1</sub>	P <sub>1</sub>	J <sub>1</sub>
۲	S <sub>1</sub>	P <sub>1</sub>	J <sub>2</sub>
۳	S <sub>1</sub>	P <sub>2</sub>	J <sub>1</sub>
۴	S <sub>2</sub>	P <sub>1</sub>	J <sub>1</sub>

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





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

□ 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.

□ 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 <sub>1</sub>	b <sub>1</sub>	c <sub>2</sub>
a <sub>1</sub>	b <sub>1</sub>	c <sub>1</sub>
a <sub>2</sub>	b <sub>1</sub>	c <sub>1</sub>

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

□ MVD ندارد.  $\Leftarrow$  4NF است.

□ CC ندارد.  $\Leftarrow$  5NF است.



بخش هفتم: طراحی پایگاه داده رابطه‌ای

رابطه R در 6NF است هر گاه در 5NF باشد و اصلاً JD نداشته باشد.



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

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

▪ با روش سنتز صفات (دسته‌بندی صفات) به تعدادی رابطه می‌رسیم.

□ در عمل :

1. ابتدا روش بالا به پایین و رسیدن به تعدادی رابطه با درجه متعارف

2. استفاده از این تئوری برای تست نرمالیتی (اول تست  $3NF$ ، بعد  $BCNF$  و  $5NF$ ) و تجزیه رابطه‌ها «در صورت نیاز»



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 نرمال گزینش 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**