

## فصل پنجم

### نرمال‌سازی

#### هدف کلی

طراحی و نرمال‌سازی رابطه‌ها، مطالعه وابستگی و مراحل نرمال‌سازی، و معرفی فرم‌های نرمال

#### هدف‌های یادگیری

۱. خوانندگان گرامی در این فصل می‌آموزند که رابطه (جدول) نرمال چیست و چگونه نرمال‌بودن آن باعث هدردادن حافظه و زمان اجرای پرسش‌ها در حد وسیعی می‌شود.
۲. مبنای اصلی نرمال‌سازی وابستگی است که انواع مختلف دارد. در حقیقت وابستگی بین ستون‌های است که آن‌ها را در یک جدول قرار می‌دهد. مثلاً در جدول شخص‌ستون‌های شماره ملی، نام، شماره شناسنامه، ... وجود دارد، ولی شماره دانشجویی جایی ندارد، زیرا هر شخصی لزوماً دانشجو نیست.
۳. انواع وابستگی، ارتباط آن‌ها با نرمال‌سازی رابطه‌ها و چگونگی کنار هم قراردادن ستون‌های مرتبط در یک جدول مطالعه می‌شود.
۴. بر مبنای انواع وابستگی، فرم‌های نرمال تعریف می‌شوند و نحوه گذار از یک فرم به فرم دیگر آموزش داده می‌شود.

## مقدمه

در طراحی پایگاهداده‌ها، بهتر است از مدل‌سازی معنایی شروع کنیم، زیرا همه چیز دورِ محور «معنا» می‌چرخد. مثلاً در مورد پایگاهداده نشر کتاب، جدول مؤلف، ویژگی‌های مربوط به نویسنده‌گان و مترجمین و ... را در برمی‌گیرد. این ویژگی‌ها، باید به‌کار موردِ نظر مربوط باشند. به عنوان نمونه، قدِ مؤلف را در جدول نمی‌آوریم، زیرا تأثیر مشخصی روی تألیف ندارد. از سوی دیگر، ویژگی‌های دیگر مربوط به تألیف، که به‌طور مستقیم به مؤلف مربوط نمی‌شود را در جدول مؤلف نمی‌آوریم. به عنوان مثال، نوع کتاب به نشر کتاب مربوط می‌شود ولی در جدول مؤلف نمی‌آید، زیرا یک مؤلف ممکن است کتاب‌هایی از انواع مختلف را تألیف کند.

به‌طور خلاصه، در طراحی جدول باید مرز استقلال معنایی را رعایت کرد، یعنی هر آنچه به جدول و کاربرد مورد نظر مربوط می‌شود، در آن باید و هیچ چیز دیگری نیاید. در تعیین ویژگی‌ها (ستون‌ها)‌ی یک جدول باید نکات زیر را در نظر داشت. این نکات را مرز استقلال معنایی می‌نامیم.

۱. همه ویژگی‌های اصلی مربوط به کاربرد مورد نظر باید. به عبارت دیگر، ویژگی یک جدول به‌طور کامل فقط در همان جدول منظور شود.

۲. ویژگی‌هایی که به‌طور مستقیم به آن جدول مربوط نمی‌شود، کنار گذاشته شود. مثلاً در طراحی جدول کارکنان، ویژگی‌هایی چون نام، نوع، تاریخ تولد، جنسیت و آدرس منظور می‌شود. ویژگی‌هایی چون ورزش مورد علاقه، دوستان و ... منظور نمی‌شود، زیرا به کاربرد مورد نظر (نشر کتاب) مربوط نیست. از سوی دیگر ویژگی‌هایی چون «کتابی که تألیف کرده‌است» در جدول کارکنان نمی‌آید، زیرا اگرچه به نشر کتاب ارتباط دارد، ولی به جدول دیگری به نام مؤلف مربوط می‌شود. جدول‌هایی که موارد بالا را رعایت نکنند، نرمال نخواهند بود و مشکلات زیادی، به‌ویژه در پایگاهداده‌های پر حجم، به وجود می‌آورند.

## ۱-۵ مشکلات جداول غیرنرمال

اگر همه ویژگی‌های مربوط به ناشر و نشر را در یک جدول منظور کنیم، این جدول به صورت زیر خواهد بود (جدول temp).

(تیراژ، قیمت واحد، تاریخ نشر، کلید کتاب، آدرس، تاریخ تأسیس، نوع انتشارات، عنوان ناشر، کلید مدیر)

این جدول سه مشکل عمده و تأثیرگذار دارد:

### ۱. افزونگی داده‌ها

فرض کنید مدیر با کلید در انتشارات الف یک میلیون کتاب چاپ کرده باشد. در این صورت برای او در انتشارات الف، یک میلیون سطر در جدول منظور می‌شود. در همه این سطراها، نوع انتشارات و تاریخ تأسیس و... تکرار می‌شود. این تکرار بسیار رویه است، یعنی می‌توان در طراحی بهتری از آن جلوگیری کرد. تکرار بسیار رویه، افزونگی داده‌ها نام دارد.

تکرار غیراجتناب، تکرار کلید است. به عبارت دیگر، برای برقراری ارتباط بین جدول‌ها، مجبوریم کلید یکی را در دیگری تکرار کنیم. این گونه تکرار کلید مجاز است، ولی تکرار سایر ویژگی‌ها، غیرمجاز است و افزونگی داده نام دارد.

سؤال: آیا افزونگی داده، امتیازی هم دارد؟

پاسخ: تنها امتیاز آن، یکجا شدن داده‌ها و عدم نیاز به پیوند جداول است، ولی در اغلب موارد، آنچنان حجم داده را افزایش می‌دهد که به هیچ وجه قابل قبول نیست. طراح می‌تواند در موارد استثنایی جدول را نرم‌ال نکند.<sup>۱</sup>

### ۲. بی‌نظمی<sup>۲</sup>

اگر طراح به ادغام جداول ادامه دهد و چند جدول را با هم بیاورد، پایگاه داده‌ها دچار سردرگمی می‌شود. این بی‌نظمی در هنگام بارگذاری پایگاه داده بیشتر مشهود است، اگر موقع وارد کردن داده‌ها، یکی از اقلام داده، اشتباه تایپ شود، نمی‌توان هویت آن را تشخیص داد. این وضعیت را بی‌نظمی می‌نامند.

مثال: در جدول temp اگر در سطراهایی نوع انتشارات الف را «کودک» و در سطراهای دیگری «کوچک» وارد کرده باشند، آنگاه این شباهه پیش می‌آید که دو انتشارات با نام الف وجود دارد.

### ۳. تکرار مقادیر تهی

بعضی از ویژگی‌های جدول، ممکن است تهی باشند. پرداختن به مقادیر تهی

1. Denormalization

2. Anomaly

زمانبر است، زیرا مقدار تهی ممکن است واقعاً تهی نباشد، ولی سیستم صلاح نداند که آن را در اختیار کاربر قرار دهد. در این صورت از هرگونه عملیات مانند عملیات جبری روی مقادیر تهی باید جلوگیری شود. حال اگر به دلیل نرمال‌بودن جداول، مقادیر تهی به صورت بی‌رویه تکرار شوند، این کنترل‌ها و بررسی‌ها هم باید تکرار شود و زمان زیادی از سیستم را هدر می‌دهد.

## ۲-۵ وابستگی و نرمال‌سازی

جدول در مرز استقلال معنایی تعریف می‌شود، یعنی آنچه در کاربرد مورد نظر به آن پدیده مربوط می‌شود، در آن منظور می‌شود و هیچ ویژگی دیگری منظور نمی‌شود. سؤال این است که مرز استقلال معنایی را چگونه می‌توان فرمول‌بندی کرد، یعنی فرایندی برای بیان و حفظ مرز استقلال معنایی داشت. پاسخ این سؤال به‌طور کلی این است: ویژگی‌هایی که به کلید یک جدول وابستگی دارند، باید در آن جدول بیانند و سایر ویژگی‌ها باید مجزا شوند.

صحت، سازگاری، دقت و اعتبار داده‌ها در شروع و پایان هر تراکنش، مهم‌ترین رکن پایگاهداده‌ها است و جامعیت نام دارد. جامعیت رابطه تنگاتنگی با وابستگی دارد. در این بخش به این مهم می‌پردازیم. عواملی که می‌توانند جامعیت را نقض کند و به وابستگی مربوط می‌شوند، عبارت‌اند از:

- وجود افزونگی داده‌ها
- اشتباهات معنایی که بر اثر عدم شناخت کافی از کاربرد مورد نظر اتفاق می‌افتد. در مرحله طراحی می‌توان با شناخت محدودیت‌های جامعیتی، آن‌ها را لحاظ کرد و در مرحله اجرا برای پیشگیری از این اتفاقات باید قیود (محدودیت‌ها) جامعیت را تعریف و اعمال کرد. یک مثال ساده، تعریف کلید رابطه‌ها است که به وابستگی مربوط است. اگر کلید رابطه‌ای را درست تعریف نکنیم، در درازمدت دچار مشکل خواهیم شد. مثلاً اگر کلید جدول کارکنان را صفت نام (name) در نظر بگیریم، ممکن است در

درازمدت افراد همنام استخدام شوند<sup>۱</sup>. از سوی دیگر اگر کلید جدول کارکنان، مثلاً شماره ملی باشد، (pno) چنین مشکلی بروز نخواهد کرد. از محدودیت جامعیتی دیگری که به مفهوم وابستگی ارتباط دارد، محدودیت ارجاع است، یعنی کلید خارجی را باید درست تعریف کرد. تعریف یک صفت به عنوان کلید خارجی، حداقل باید دو شرط زیر را برآورده سازد که به آن‌ها کلان‌قید می‌گویند.

۱. آن صفت در جدول مرجع کلید باشد. مثلاً تعریف bno در جدولی به عنوان کلید خارجی که به جدول book ارجاع دارد صحیح نیست، زیرا bno به تنها یکی در جدول book کلید نیست (bno و ano با هم کلید هستند).

البته می‌توان از bno در هرجا برای شماره یک کتاب بدون توجه به مؤلف آن استفاده کرد.

۲. مقدار آن صفت باید در جدول مرجع وجود داشته باشد. مثلاً اگر در جدول ناشر، مدیر نشری با کلید ۱۱ بیاید، حتماً باید شخصی با کلید ۱۱ در جدول کارکنان نیز وجود داشته باشد.

این قیود جامعیت که به وابستگی مربوط می‌شود، کلان‌قید یا قیود ساختاری<sup>۲</sup> نام دارد زیرا به کاربر محدود نمی‌شود، یعنی در هر پایگاه‌داده‌ای باید اعمال شوند [Date 2003]. در صفحات آینده درباره اهمیت انواع کلید و تأثیر آن در وابستگی و نرمال‌سازی جداول بحث می‌کنیم. تشخیص صحیح کلیدهای کاندید و کلید اصلی و کلیدهای خارجی جداول اهمیت ویژه‌ای دارد که الگوریتمی برای آن ارائه خواهیم داد.

### ۳-۵ وابستگی تابعی<sup>۳</sup>

کلید جدول، مجموعه‌ای از ویژگی‌ها (صفتها) است که سایر صفت‌ها را می‌توان از آن به دست آورد. مثلاً کد ملی یک فرد، کلید او است، یعنی در هر جدول نرمالی می‌توان مشخصات موجود آن فرد را از روی کد ملی او به دست آورد. حالت کلی این مفهوم را وابستگی تابعی می‌نامیم و با → نمایش می‌دهیم.

۱. اصولاً کلیدهای اصلی جداول بهتر است عددی باشند، زیرا معمولاً در جداول دیگر به صورت کلید خارجی می‌آیند و علاوه بر تحمیل حجم اضافی، کنترل‌های لازم روی صفت‌های غیر عددی مانند متغیر بودن طول آن‌ها، مشکل آفرین است.

2. Structural Constraint  
3. Functional Dependency

کد ملی → سایر ویژگی‌های فرد

مثال: در پایگاهداده نشر کتاب در جدول کارکنان

$Pno \rightarrow name, typ, dob, sex, adres$

این وابستگی تابعی را می‌توان جداگانه نیز نوشت، به صورت زیر:

$Pno \rightarrow name$        $Pno \rightarrow dob$        $Pno \rightarrow typ$   
 $Pno \rightarrow sex$        $Pno \rightarrow adres$

وابستگی تابعی به کلید محدود نمی‌شود، یعنی در یک جدول (معمولًاً غیرنرم‌ال)

ممکن است وابستگی‌های تابعی دیگری نیز وجود داشته باشد.

مثال: در جدول temp علاوه بر وابستگی به کلید، وابستگی‌های زیر را نیز داریم:

تاریخ نشر → کلید کتاب

تیراز و قیمت واحد → (کلید کتاب، تاریخ نشر)

هدف اصلی نرم‌ال‌سازی جدول، این است که وابستگی به کلید را حفظ کنیم و سایر وابستگی‌ها را با شکستن یک جدول و تبدیل آن به چند جدول، به وابستگی به کلید تبدیل کنیم.

مثال: جدول زیر را در نظر بگیرید:

$T(A, B, C, D)$

در این جدول A کلید است، یعنی  $A \rightarrow B, C, D$  ولی وابستگی تابعی اضافی  $C \rightarrow D$  نیز وجود دارد. پس این جدول نرم‌ال نیست. بعداً خواهیم دید که باید آن را به دو جدول  $T_1$ ،  $T_2$  زیر تبدیل کرد که معادل جدول T هستند:

$T_1(A, B, C)$

$T_2(C, D)$

علاوه بر وابستگی تابعی، وابستگی‌های دیگری نیز بین صفت‌های یک جدول می‌تواند وجود داشته باشد، مانند وابستگی چندمقداری و وابستگی پیوندی. در ادامه به هر سه مورد می‌پردازیم.

### ۱-۳-۵ تعاریف

مفاهیم و تعاریف زیر از همان آغاز پیدایش مدل رابطه‌ای، توسط آقای کاد ارائه شده‌اند

: [ 1970، E. F. Codd]

تعریف:

اگر  $A \rightarrow B$  دو مجموعه صفت در شمای  $R$  باشند، آنگاه وابستگی تابعی  $A \rightarrow B$  برقرار است اگر، برای تمام رابطه‌ها در  $R$ ، به ازای هر مقدار  $A$  فقط یک مقدار  $B$  داشته باشیم.

در این تعریف نکات زیر قابل توجه است:

۱. وابستگی تابعی باید برای تمام رابطه‌ها درست باشد، یعنی از معنی و مفهوم آن صفت‌ها سرچشمme بگیرد، نه از موارد خاص در یک یا چند رابطه.
۲. تأکید می‌کنیم که از وابستگی می‌توان برای تعریف قید جامعیت نیز استفاده کرد.

تعریف:

$A \rightarrow B$  را وابستگی تابعی کامل<sup>۱</sup> می‌گوییم اگر  $B$  به هیچ زیرمجموعه مخصوص<sup>۲</sup> از  $A$  وابسته نباشد.

تعریف:

اگر برای تمام صفت‌های  $B$  در جدول  $R$  داشته باشیم  $A \rightarrow B$  آنگاه  $A$  را «ابرکلید  $R$ » می‌نامند و به صورت  $A \rightarrow R$  نمایش می‌دهند.

اگر این وابستگی از نوع FFD باشد، آنگاه  $A$  کلید کاندید  $R$  است.

تعریف:

اگر  $B$  زیرمجموعه‌ای از  $A$  باشد آنگاه همواره داریم  $B \rightarrow A$ . این نوع وابستگی تابعی را، بدیهی<sup>۳</sup> می‌نامیم.

مثال: در جدول book داریم:

bno ↗ title, dop, typ, version

bno ↗ ano

اما:

زیرا یک کتاب ممکن است چند مؤلف داشته باشد. مثلاً اگر شماره کتاب پایگاه‌داده‌ها  $x$  باشد، آنگاه

حق‌جو ↗  $x$

فراهی ↗  $x$

1. Full Functional Dependency یا FFD

2. Proper Subset

3. Trivial FD

بنابراین bno به تنها یک کلید جدول کتاب نیست، زیرا همه صفت‌های آن را نمی‌دهد. (bno ano)، با هم کلید کاندید این جدول هستند، زیرا علاوه بر کلیدبودن، کمینه نیز هست، یعنی: (bno,ano) → book

ولی اگر ستون‌های دیگری به این مجموعه اضافه کنیم، باز هم کلید است. مثلاً:

(bno,ano,title) → book

این کلید، دیگر کمینه نیست و نمی‌تواند کلید کاندید باشد.

نتیجه می‌گیریم که:

۱. همه وابستگی‌های تابعی در یک جدول اهمیت دارند. اگر یک یا چند وابستگی از قلم بیفتند، نرمال‌سازی به درستی انجام نمی‌شود.

۲. بعضی از وابستگی‌ها، اضافی یا تکراری یا بی‌اثر هستند، مانند  
 $bno \rightarrow bno$   
 $(bno,title) \rightarrow title$

۳. کلیدهای کاندید جدول، جایگاه ویژه‌ای در نرمال‌سازی آن دارند. پس برای نرمال‌سازی جدول، ابتدا باید همه وابستگی‌های آن را بیابیم تا به مجموعه پوششی وابستگی بررسیم، سپس زوائد آن را دور بریزیم تا به مجموعه بهینه وابستگی بررسیم. در پایان، کلیدهای کاندید را بیابیم و آنگاه اقدام به نرمال‌سازی کنیم [حق‌جو، ۱۳۹۲].

تعریف:

اگر F یک مجموعه از وابستگی‌های تابعی باشد، آنگاه مجموعه تمام وابستگی‌های تابعی که از آن نتیجه می‌شود را مجموعه پوششی F می‌نامیم و با F+ نمایش می‌دهیم.

### ۲-۳-۵ یافتن مجموعه پوششی و مجموعه بهینه وابستگی

آقای آرمسترانگ در سال ۱۹۷۴ روشی برای استخراج مجموعه پوششی ارائه داده است [W. Armstrong, 1974]. او ثابت کرد که سه قاعدة زیر برای استخراج مجموعه پوششی کافی است، یعنی با اعمال مکرر این سه قاعدة می‌توان به تمام وابستگی‌های نتیجه‌شده دست یافت:

۱. بازتاب<sup>۱</sup>: اگر B زیرمجموعه A باشد، آنگاه  $A \rightarrow B$

۲. افزایش<sup>۱</sup>: اگر  $A \rightarrow B$  و  $C$  صفت باشد، آنگاه  $AC \rightarrow BC$

۳. انتقال<sup>۲</sup>: اگر  $A \rightarrow C$  و  $B \rightarrow A$ ، آنگاه  $B \rightarrow C$

هرچند قاعده‌های فوق برای استخراج  $F +$  کفايت می‌کرد، ولی اعمال آنها مشکل بود. بعدها، دیگران قواعد دیگری را بیان کردند که کار را سهولت بخشید. مهم‌ترین آنها عبارت‌اند از:

۱. اجتماع<sup>۳</sup>: اگر  $A \rightarrow BC$  و  $A \rightarrow C$ ، آنگاه  $A \rightarrow B$

۲. تجزیه<sup>۴</sup>: اگر  $A \rightarrow BC$ ، آنگاه  $A \rightarrow B$  و  $A \rightarrow C$

۳. ترکیب<sup>۵</sup>: اگر  $AC \rightarrow BD$  و  $C \rightarrow D$ ، آنگاه  $A \rightarrow B$

با اعمال قواعد فوق، وابستگی‌های زیادی به دست می‌آید که بعضًا تکراری و اضافی هستند. در اینجا روشی برای حذف این‌گونه وابستگی‌ها و رسیدن به مجموعه وابستگی بهینه ارائه می‌دهیم.

تعریف:

دو مجموع G وابستگی تابعی  $F_1$  و  $F_2$  را معادل<sup>۶</sup> می‌نامیم، اگر مجموعه پوششی آنها برابر باشد، یعنی  $F_1+ = F_2 +$

با استفاده از قواعد سه‌گانه زیر می‌توان یک مجموعه وابستگی را به مجموعه بهینه معادل آن تبدیل کرد (Fopt):

۱. سمت راست هر وابستگی فقط یک صفت باشد.

۲. هر صفتی که  $+F$  را تغییر نمی‌دهد، از سمت چپ حذف شود.

۳. وابستگی‌های تکراری و اضافی حذف شود.

برای یافتن وابستگی‌های تابعی در پایگاهداده، می‌بایست مجموعه پوششی وابستگی‌ها را تعیین و آن را بهینه کرد. می‌توان ابتدا F را ساده کرد و سپس یافتن  $+F$  و Fopt را همزمان انجام داد.

یادآوری: در مدل شی-رابطه‌ای برای جداول تودرتو، باید نرمالسازی را

1. Augmentation
2. Transitivity
3. Union
4. Decomposition
5. Composition
6. Equivalent

به صورت پایین به بالا<sup>۱</sup> انجام داد، یعنی ابتدا جداول داخلی را نرمال کرد، سپس آن‌ها را کنار گذاشت و ستون‌های ساده جداول بیرونی را نرمال کرد. در پایان، جداول معادل، جایگزین هریک از جداول داخلی می‌شوند.

مثال: در بانک اطلاعات زیر، مجموعه وابستگی پوششی بهینه را بیابید:

$$R = (U, V, W, X, Y, Z)$$

$$F = \{U \rightarrow XY, X \rightarrow Y, XY \rightarrow ZV\}$$

حل: ابتدا  $F^+$  را می‌یابیم و سپس آن را بهینه می‌کنیم.

$$F^+ = \{U \rightarrow XY, X \rightarrow Y, XY \rightarrow ZV, U \rightarrow ZV, X \rightarrow ZV\}$$

قاعده ۱ روی وابستگی اول:

$$F^+ = \{U \rightarrow X, U \rightarrow Y, X \rightarrow Y, XY \rightarrow ZV, U \rightarrow ZV, X \rightarrow ZV\}$$

قاعده ۲ روی  $XY \rightarrow ZV$  با توجه به  $X \rightarrow Y$

$$F^+ = \{U \rightarrow X, U \rightarrow Y, X \rightarrow Y, X \rightarrow ZV, U \rightarrow ZV\}$$

حال قاعده ۱ و ۳ روی سایر وابستگی‌ها:

$$F_{opt} = \{U \rightarrow X, U \rightarrow Y, U \rightarrow Z, U \rightarrow V, X \rightarrow Y, X \rightarrow Z, X \rightarrow V\}$$

### ۳-۳-۵ کلیدهای کاندید

اگر مجموعه‌ای از صفت‌ها را ATTR و مجموعه وابستگی تابعی آن‌ها را F بنامیم (ترجیحاً بهینه)، آنگاه الگوریتم زیر، مجموعه تمام صفت‌های وابسته به ATTR را می‌دهد:

$$ATTR^+ = ATTR$$

تکرار کن

برای هر  $X \rightarrow Y$  در F

اگر X زیرمجموعه ATTR<sup>+</sup> باشد آنگاه

$$ATTR^+ = ATTR^+ \cup UNION Y$$

تا زمانی که ATTR<sup>+</sup> دیگر تغییر نکند.

با استفاده از این الگوریتم می‌توان کلیدهای کاندید را به دست آورد. بانک اطلاعات ممکن است چند کلید کاندید داشته باشد. کلیدهای کاندید ممکن است در یک یا چند صفت مشترک باشند. این نکات چنانکه خواهیم دید در نرمال‌سازی اهمیت ویژه‌ای دارند و به تجزیه یک جدول به جداول نرمال کمک می‌کنند.

مثال: اگر

$$R = (S, T, U, V, W, X, Y)$$

$$F = \{S \rightarrow T, V \rightarrow SW, T \rightarrow U, SX \rightarrow Y\}$$

آنگاه

$$S^+ = \{S, T, U\}$$

$$V^+ = \{V, S, W, T, U\}$$

$$SX^+ = \{S, X, T, U, Y\}$$

برای یافتن کلیدها از  $S^+$  و  $V^+$  و  $SX^+$ ، می‌توان از هر کدام که بیشترین تعداد صفت‌ها را در بر می‌گیرند، شروع کرد. اگر از  $V^+$  شروع کنیم، خواهیم داشت  $.VXY \rightarrow R$ .

اما چون کلید کاندید باید کمینه باشد، پس  $VX \rightarrow R$  (زیرا  $VX \rightarrow S$   $SX \rightarrow Y$ )  $VX \rightarrow R$  است. اگر از  $SX^+$  شروع کنیم خواهیم داشت  $SXW \rightarrow R$ . پس از کمینه کردن،  $S$  و  $W$  حذف می‌شوند و خواهیم داشت  $VX \rightarrow R$  که همان کلید قبلی است و کلید کاندید دیگری به دست نمی‌آید.

### ۴-۳-۵ نمودار وابستگی تابعی<sup>۱</sup>

ابتدا باید مجموعهٔ بهینهٔ وابستگی‌ها را به دست آوریم و آنگاه به رسم نمودار وابستگی تابعی مبادرت کنیم. در این نمودار، صفت‌های ترکیبی در مستطیل قرار می‌گیرند. پیکانی از صفت‌ها به هریک از صفت‌های وابسته به آن‌ها رسم می‌شود. برای مجموعهٔ صفت‌ها و وابستگان آن‌ها نیز مستطیل‌های دیگر و پیکان‌های دیگری رسم می‌شود.

مثال: در بانک اطلاعات زیر کلیدهای کاندید را بیابید و نمودار وابستگی را رسم کنید.

$$R = (S, T, U, V, X, Z, O, P, Q)$$

$$F = \{U \rightarrow VXQ, UVQ \rightarrow O, OQ \rightarrow Z, UP \rightarrow X, P \rightarrow ST\}$$

حل: ابتدا مجموعهٔ بهینهٔ معادل  $F$  را پیدا می‌کنیم (تمرین):

$$F_{opt} = \{U \rightarrow V, U \rightarrow X, U \rightarrow Q, P \rightarrow S, P \rightarrow T, OQ \rightarrow Z, UP \rightarrow O, UP \rightarrow Z\}$$

سپس مجموعهٔ صفت‌های وابسته به تمام مجموعهٔ صفت‌های چپ پیکان‌ها را می‌یابیم:

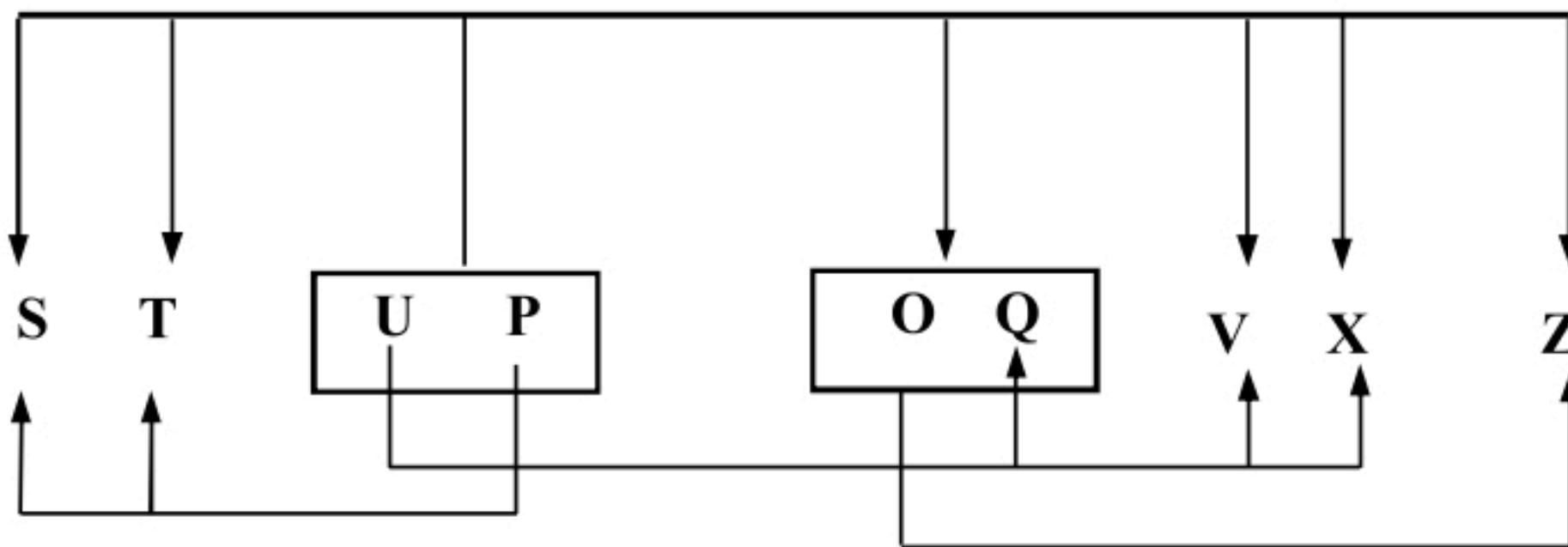
$$U+ = \{U\} \Rightarrow \{U, V, X, Q\}$$

$$P+ = \{P\} \Rightarrow \{P, S, T\}$$

$$UP+ = \{U, P\} \Rightarrow \{U, P, V, X, Q, O, S, T\} \Rightarrow \\ \{S, TU, P, V, X, O, Q, Z\}$$

$$OQ+ = \{O, Q\} \Rightarrow \{O, Q, Z\}$$

نتیجهٔ می‌گیریم که  $UP \rightarrow R$  کلید کاندید می‌باشد، یعنی  $UP \rightarrow R$  کلید کاندید دیگری نیست، زیرا از  $OQ$  نمی‌توان به  $U$  و  $P$  رسید. نمودار وابستگی این بانک اطلاعات در شکل ۱-۵ مشاهده می‌شود.



شکل ۱-۵. نمودار وابستگی تابعی

مثال: در پایگاه‌داده زیر:

۱. مجموعهٔ وابستگی بهینه را بیابید.

۲. نمودار وابستگی تابعی را رسم کنید.

$$R = (A, B, C, D, E, F, G)$$

$$F = \{(A, B) \rightarrow R, A \rightarrow G, B \rightarrow EF, G \rightarrow DF\}$$

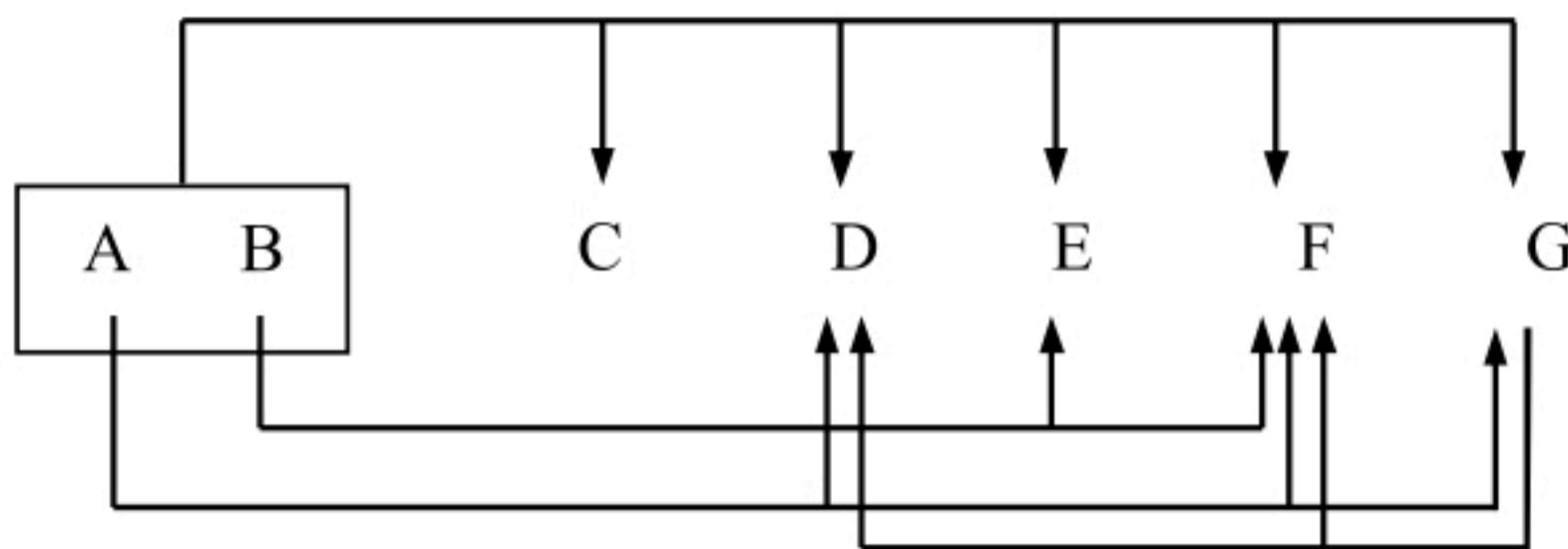
حل: باید وابستگی  $F \cup A \rightarrow DF$  اضافه شود یعنی  $A \rightarrow DF$

$$F_{opt} = \{(A, B) \rightarrow R, A \rightarrow G, B \rightarrow E, B \rightarrow F, G \rightarrow D, G \rightarrow F, A \rightarrow D, A \rightarrow F\}$$

پس از مرتب‌کردن صفت‌ها:

$$F_{opt} = \{(A, B) \rightarrow R, A \rightarrow D, A \rightarrow F, A \rightarrow G, B \rightarrow E, B \rightarrow F, G \rightarrow D, G \rightarrow F\}$$

نمودار وابستگی تابعی به دست آمده در شکل ۲-۵ مشاهده می‌شود:



شکل ۲-۵. نمودار وابستگی تابعی دیگر

#### ۴-۵ تجزیه جداول به فرم‌های نرمال در وابستگی تابعی

آقای کاد در همان مقاله اولیه خود سه فرم نرمال را معرفی کرد که به 1NF و 2NF و 3NF معروف هستند<sup>۱</sup>. چیزی نگذشت که دانشجوی دکتری ایشان به نام بویس همراه با ایشان، فرم نرمال دیگری به نام BCNF تعریف کرد [J. Biskup 1979]. این چهار فرم نرمال بر مبنای وابستگی تابعی تعریف شده‌اند. بعدها دیگران وابستگی‌ها و فرم‌های نرمال دیگری مانند 4NF و 5NF و... را معرفی کردند که نادر هستند و در موارد ویژه‌ای کاربرد دارند. در این فصل، نرمال‌سازی تا سطح 5NF بررسی می‌شود.

فرم نرمال 1NF اصلی‌ترین مبنای مدل رابطه‌ای را بیان می‌کند. فرم‌های نرمال دیگر، هر کدام مشکلی را که در فرم قبلی وجود دارد، حل می‌کنند. هرچند فرم 1NF به مدل رابطه‌ای غنا می‌بخشد، اما آن را نیز محدود می‌کند. در این بخش فرم‌های اصلی نرمال ارائه می‌شود.

#### ۱-۴-۵ فرم نرمال اول

تعریف: جدولی در 1NF است که:

۱. همه کلیدهای کاندید آن تعریف شده باشند.

۲. تمام صفت‌های آن به کلید اصلی، وابستگی تابعی داشته باشند.

۳. صفت‌های آن از دامنه تودرتو<sup>۲</sup> نباشند.

منظور از شرط سوم این است که صفات ترکیبی نداشته باشیم، یعنی مثلاً یا صفتی به نام آدرس بدون توجه به بخش‌های آن (شهر و...) یا بخش‌های آن را جداگانه بدون توجه به کل داشته باشیم.

1. First Normal Form, Second Normal Form and Third Normal Form  
2. Nested Domain

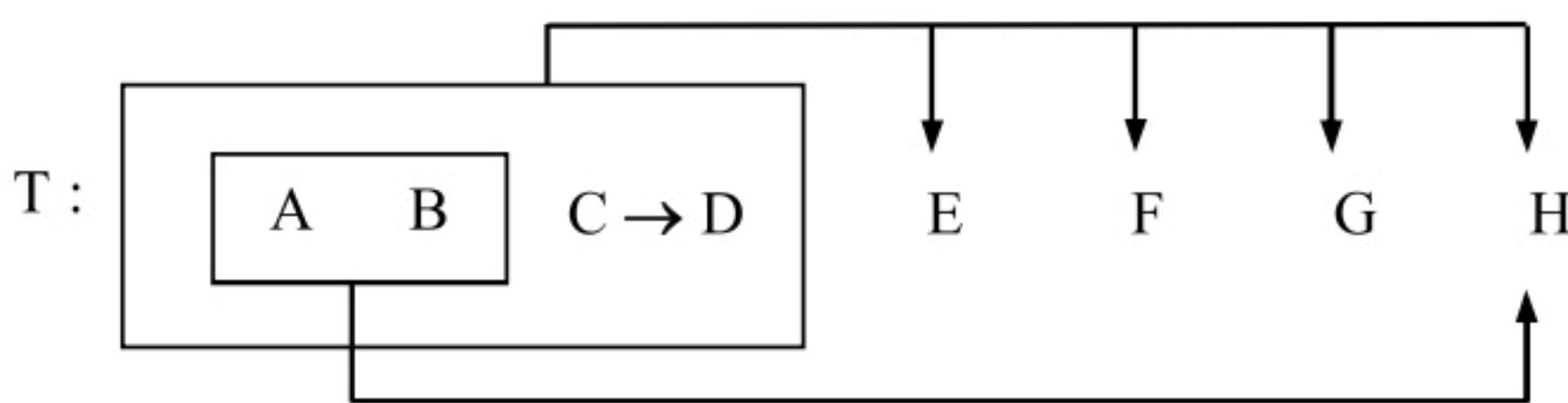
### ۲-۴-۵ فرم نرمال دوم

تعریف: جدولی در فرم نرمال دوم ( $2NF$ ) است که:

۱. در  $1NF$  باشد.
۲. وابستگی جزئی<sup>۱</sup> نداشته باشد، یعنی صفت‌های آن به زیرمجموعه‌های کلید اصلی وابستگی نداشته باشند.

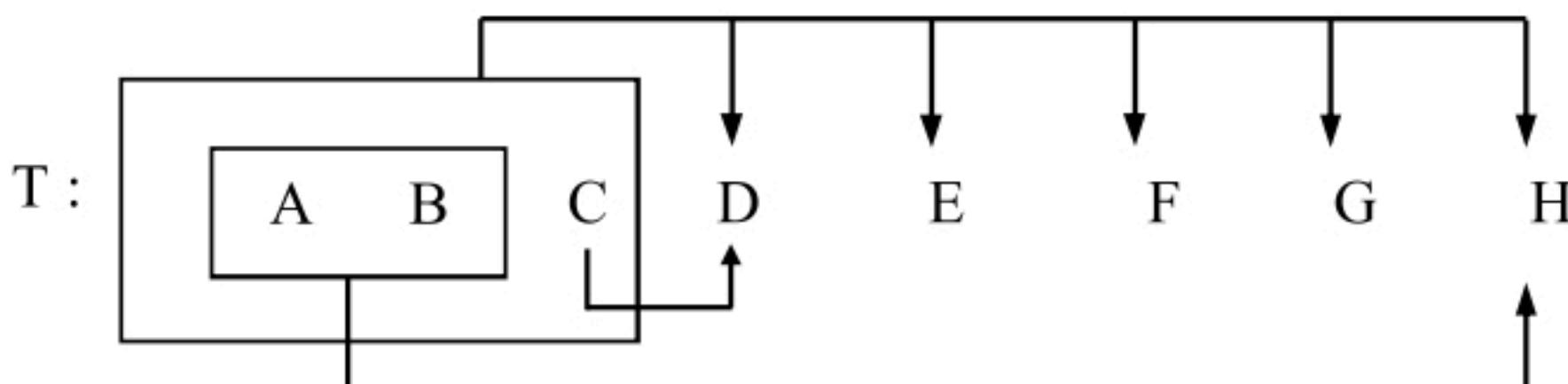
الگوریتم زیر، جدول  $1NF$  را به جداول  $2NF$  تجزیه می‌کند:

۱. هر بخش از کلید اصلی را که ایجاد وابستگی جزئی کرده است، با همه وابسته‌های آن کنار هم بگذار.
  ۲. کل کلید اصلی را با صفت‌های باقی‌مانده کنار هم بگذار.
  ۳. صفت‌های کلیدی را به عنوان کلید خارجی به ۲ اضافه کن.
- مثال: نمودار وابستگی تابعی در شکل ۳-۵ را نرمال کنید.



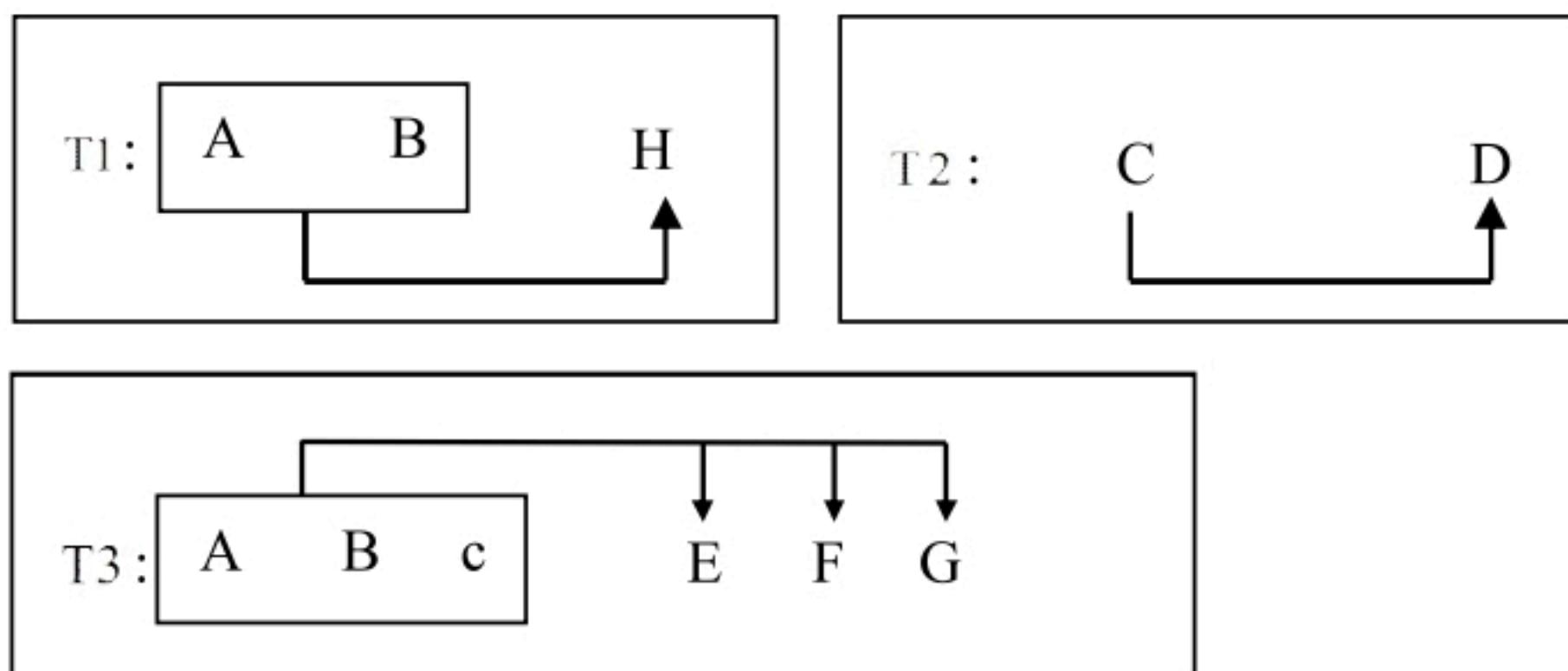
شکل ۳-۵. مثال برای فرم‌های نرمال اول و دوم

حل: این نمودار غلط است، زیرا کلید آن کمینه نیست؛ به عبارت دیگر این نمودار در فرم  $1NF$  نیست. نمودار صحیح به صورت شکل ۴-۵ است.



شکل ۴-۵. نمودار اصلاح شده فرم نرمال دوم

پیروی از الگوریتم تبدیل از فرم نرمال اول به دوم، شکل زیر را که در فرم نرمال دوم است، به دست می‌دهد. شکل نرمال این پایگاهداده شامل جداول T1 و T2 و T3 است که معادل جدول T هستند.



### ۳-۴-۵ صحت تجزیه جدول

#### ۱. قاعدة پیوندپذیری<sup>۱</sup>

در تجزیه جدول، باید از پیوند طبیعی جداول‌های تجزیه شده، دقیقاً جدول اصلی حاصل شود [A. V. Aho, 1979] و [J. Biskup, 1979]

تعریف:

تجزیه شمای R به  $\{R_1, \dots, R_n\}$ . در صورتی «پیوندپذیر» است که برای تمام جداول r از شمای R داشته باشیم:

$$r = \prod_{R_1}(r) \times \dots \times \prod_{R_n}(r)$$

این تعریف و همه تعاریف مشابه با توجه به معنی داده‌ها و صفت‌ها روی شمای جداول بیان می‌شود، یعنی باید برای تمام جداول درست باشد.

واضح است که این تعریف قابل پیاده‌سازی نیست، زیرا نمی‌توان تمام جداول ممکن روی یک شما را بررسی کرد. پژوهش‌گران راه حل ساده و عملی زیر را یافته‌اند: «تجزیه R به  $\{R_1, R_2\}$  در صورتی پیوندپذیر است که کلید حداقل یکی از دو جدول، ستون‌های مشترک آن دو باشد»

به عبارت دیگر:

$$\{R = \{R1, R2\}\} \Rightarrow \{R1 \cap R2 \rightarrow R1 \text{ یا } R1 \cap R2 \rightarrow R2\}$$

این قاعده برای دو جدول بیان شده است. در مواردی که جدولی به بیش از دو جدول تجزیه شود، باید صحت آن را قدم به قدم بررسی کرد.

## ۲. قاعده حفظ وابستگی

تجزیه یک شما به چند شمای کوچک‌تر باید «حافظ وابستگی<sup>۱</sup>» باشد. یعنی تمامی وابستگی‌های اصلی حفظ شود. پژوهش‌گران راه عملی برای بررسی «حافظ وابستگی بودن» یک تجزیه را نیز پیدا کرده‌اند. این راه حل به صورت زیر است: [A. V. Aho, 1979] و [J. Biskup 1979]

یادآوری: هر بانک اطلاعات رابطه‌ای را همواره می‌توان به صورت مجموعه  $\{R, F\}$ ، نمایش داد، یعنی شمای  $R$  با مجموعه وابستگی  $F$  تجزیه بانک اطلاعات  $\{R, F\}$  به  $\{R1, F1\}, \dots, \{Rn, Fn\}$  در صورتی حافظ وابستگی است که داشته باشیم:

$$F+ = \{F1 \dot{\cup} \dots \dot{\cup} Fn\} +$$

## ۴-۴-۵ فرم نرمال سوم

تعریف: جدولی در 3NF است که:

۱. در 2NF باشد.

۲. وابستگی انتقالی<sup>۲</sup> (وابستگی صفت‌های غیرکلیدی به یکدیگر) نداشته باشد.

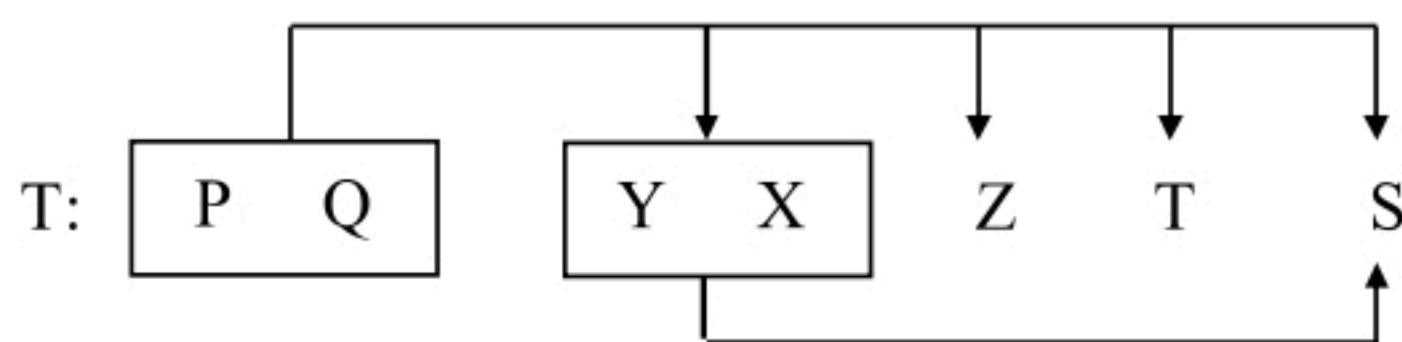
الگوریتم زیر، تبدیل از 2NF به 3NF را بیان می‌کند:

۱. مجموعه صفاتی را که وابستگی انتقالی ایجاد کرده‌اند، با همه وابسته‌های آن‌ها کنار هم بگذار،

۲. کلید اصلی را با صفت‌های باقی‌مانده کنار هم بگذار،

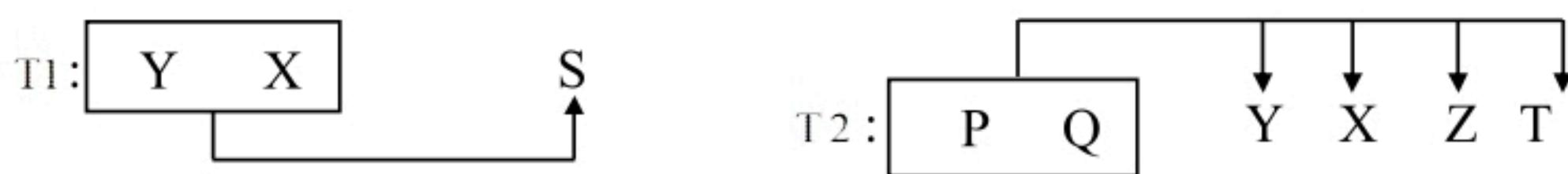
۳. صفت‌های کلیدی را به عنوان کلید خارجی به ۲ اضافه کن.

مثال: نمودار وابستگی تابعی شکل ۵-۵ را نرمال کنید



شکل ۵-۵. مثال برای فرم نرمال سوم

حل: این نمودار در فرم نرمال اول و دوم است، زیرا کلید آن به درستی تعریف شده و وابستگی جزئی ندارد، اما در فرم نرمال سوم نیست، زیرا وابستگی انتقالی دارد یعنی صفت  $S$  به  $(Y, X)$  وابسته است و هیچ‌کدام از سه صفت، کلیدی نیستند. با پیروی از الگوریتم تبدیل به فرم نرمال سوم به دو جدول  $T_1$  و  $T_2$  به صورت زیر تجزیه می‌شود.



قاعده پیوندپذیری رعایت شده است، زیرا از پیوند طبیعی دو جدول به جدول اصلی می‌رسیم.

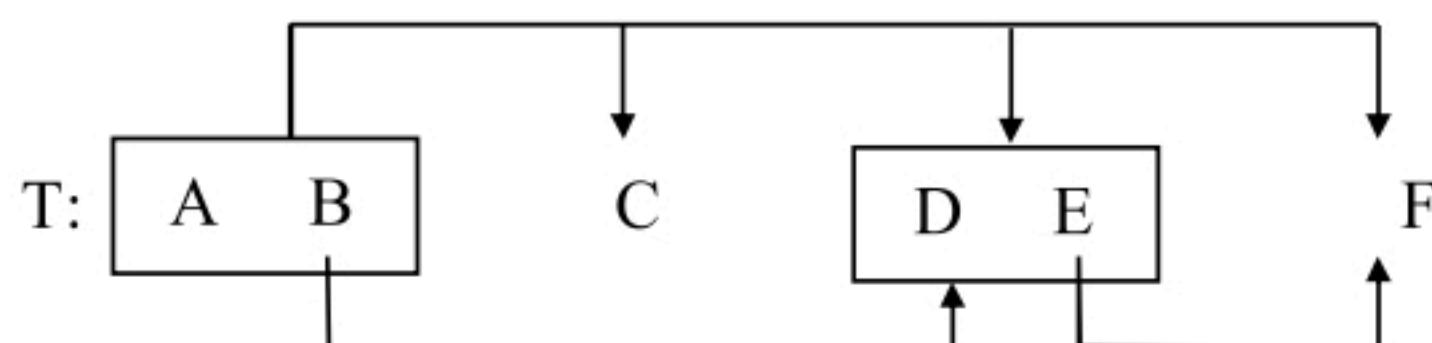
قاعده حفظ وابستگی نیز رعایت شده است، زیرا:

$$F + (T_1, T_2) = \{(Y, X) \rightarrow S, (P, Q) \rightarrow YXZT, (P, Q) \rightarrow S\}$$

$$F + (T_1, T_2) = F + (T)$$

نکته: در  $T_2$  صفت‌های  $X$ ,  $Y$  در مستطیل قرار نمی‌گیرند، زیرا سمت چپ وابستگی نیستند.

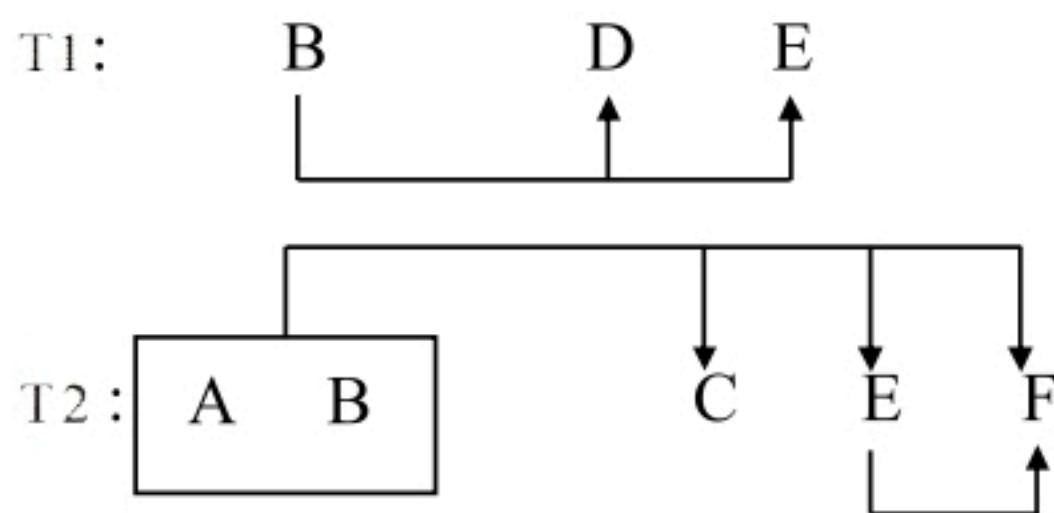
مثال ترکیبی: نمودار وابستگی تابعی شکل ۵-۶ را نرمال کنید.



شکل ۵-۶. مثال ترکیبی برای فرم‌های نرمال دوم و سوم

حل: این نمودار در فرم نرمال اول است، چون کلید آن به درستی تعیین شده ولی در فرم‌های نرمال دوم و سوم نیست. ابتدا باید آن را به فرم نرمال دوم تبدیل کنیم:

#### ۱. تبدیل به 2NF

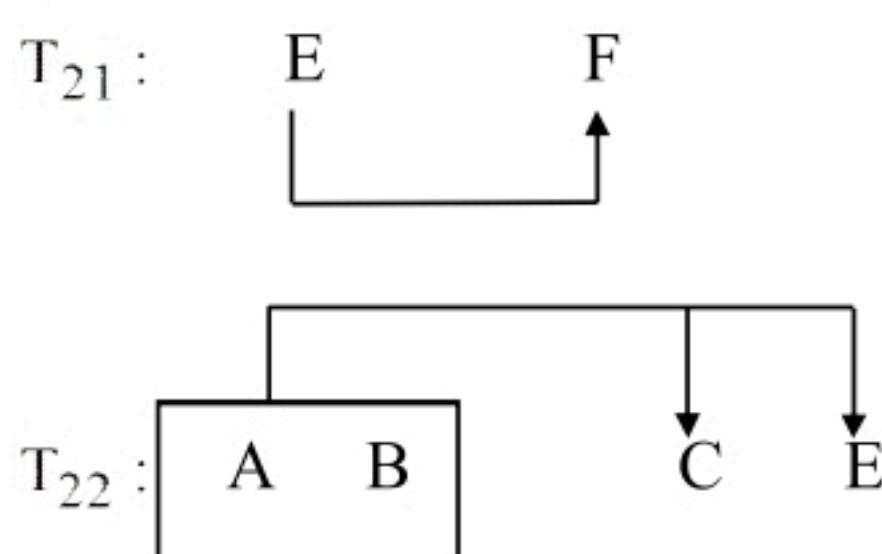


نکته:

- ستون E در T2 تکرار شده زیرا طبق قاعد، حفظ وابستگی، ستون‌های E و F باید در یک جدول باشند و ستون E نمی‌تواند در T1 بیاید زیرا  $F \not\rightarrow B$ .
- ستون‌های D و E در جدول T1 در مستطیل قرار نمی‌گیرند، زیرا سمت چپ وابستگی نیستند.

#### ۲. تبدیل به 3NF

جدول T1 در 3NF هست و جدول T2 به دو جدول زیر تجزیه می‌شود:



نتیجه: جدول T به سه جدول T1 و T<sub>21</sub> و T<sub>22</sub> تجزیه می‌شود که هر سه فقط وابستگی به کلید دارند و قواعد تجزیه جدول نیز رعایت شده است.

$$F + (T) = F + (T_1, T_{21}, T_{22})$$

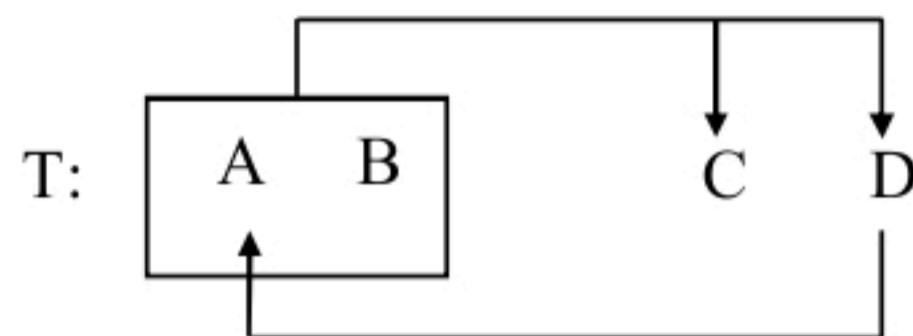
### ۴-۵-۵ فرم نرمال BCNF

دانشجوی دکتری آقای کاد به نام بویس دریافت که اگر جدولی تمام شرط‌های فرم‌های 1NF و 2NF و 3NF را داشته باشد، باز هم ممکن است وابستگی مزاحم دیگری

داشته باشد و درنتیجه، نرمال نباشد. این وابستگی مزاحم از نوع وابستگی کلید به صفت‌های غیرکلیدی یا وابستگی وارون است.

برای حل این مشکل، استاد و شاگرد با یکدیگر فرم نرمال<sup>۱</sup> BCNF را پیشنهاد کردند.

مثال: آیا جدولی با نمودار وابستگی تابعی شکل ۷-۵ نرمال است؟



شکل ۷-۵. مثال برای فرم نرمال BCNF

حل:

۱. در فرم نرمال 1NF هست، زیرا کلیدش مشخص شده است.

۲. در فرم نرمال 2NF هست، زیرا وابستگی جزیی ندارد.

۳. در فرم نرمال 3NF هست، زیرا وابستگی انتقالی ندارد.

با این همه، این جدول نرمال نیست، زیرا علاوه بر وابستگی به کلید، وابستگی اضافی  $A \rightarrow D$  را نیز دارد (وابستگی وارون).

تعریف:

جدولی در BCNF است که ستون‌های آن فقط به کلیدهای کاندیدش وابستگی تابعی داشته باشند.

توجه به نکات زیر در مورد BCNF ضروری است:

- برخلاف فرم‌های نرمال دیگر، BCNF بدون استفاده از 3NF و فرم‌های قبلی نرمال تعریف می‌شود. غالباً می‌توان بانک اطلاعات را با استفاده از تعریف BCNF نرمال کرد و نیازی به تعریف وابستگی‌های دیگر نیست.

بنابراین، BCNF جامع‌ترین تعریف نرمال‌سازی برمنای وابستگی تابعی را به‌طور مستقل ارائه می‌دهد.

- در مواردی، نرمال‌سازی تا سطح BCNF لازم نیست و بهتر است از تبدیل جدول فرم 3NF به BCNF خودداری کرد. به عنوان مثال، برای ثبت آدرس فعلی و آدرس

قبلی کارکنان می‌توان جدول زیر را تعریف کرد و برای هریک از کارکنان بیش از یک آدرس در نظر گرفت:

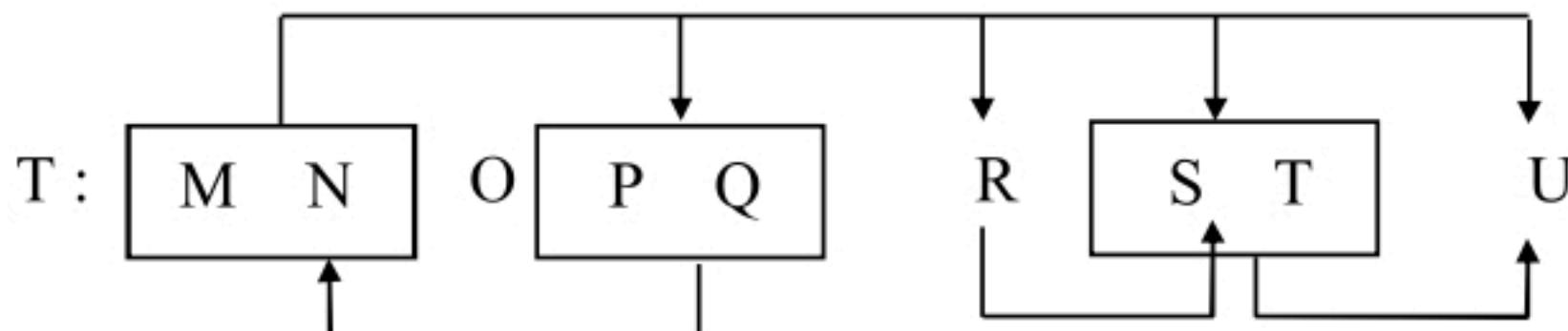
p-addr (pno, city, no, zip)  
candidate key (pno, city)  
candidate key (pno, zip)

کدپستی برای هر شهر منحصر به فرد است و می‌تواند به جای نام شهر قرار گیرد.

این جدول در فرم 3NF هست ولی در فرم BCNF نیست، زیرا داریم:  $\text{zip} \rightarrow \text{city}$   
با این همه، نمی‌توان با اطمینان خاطر این جدول را شکست، زیرا کدپستی،  
بخش جدایی‌ناپذیر آدرس است و جداکردن آن باعث پیچیده شدن پرس‌وجوهای  
مربوط به آدرس می‌شود.

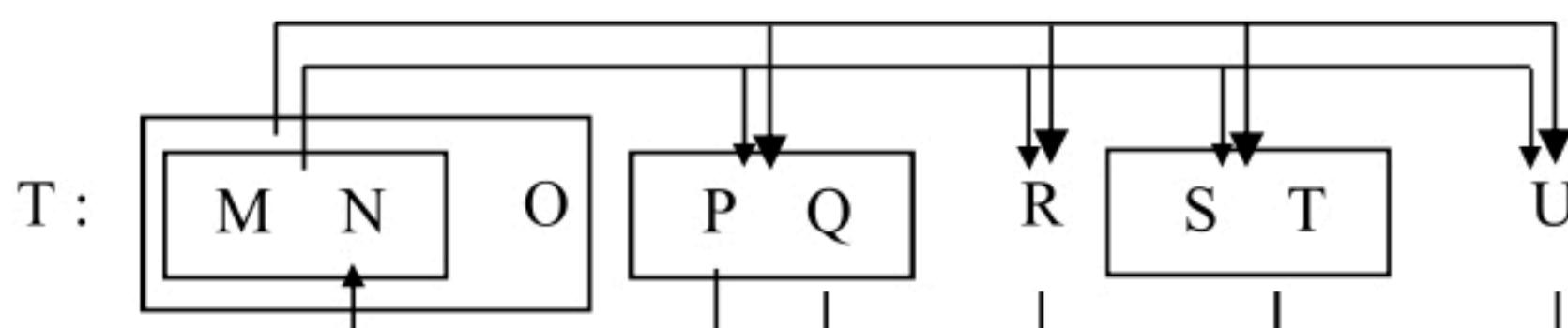
مجموعه پرسش‌ها نیز می‌تواند در طراحی جداول نقش داشته باشد. اگر طراح تشخیص دهد که تجزیه یک جدول، هرچند افزونگی هم داشته باشد، باعث پایین‌آمدن سرعت پرسش‌ها می‌شود، مجاز است از نرمال‌سازی بیشتر آن صرف‌نظر کند.

مثال ترکیبی: نمودار وابستگی شکل ۸-۵ را نرمال کنید.



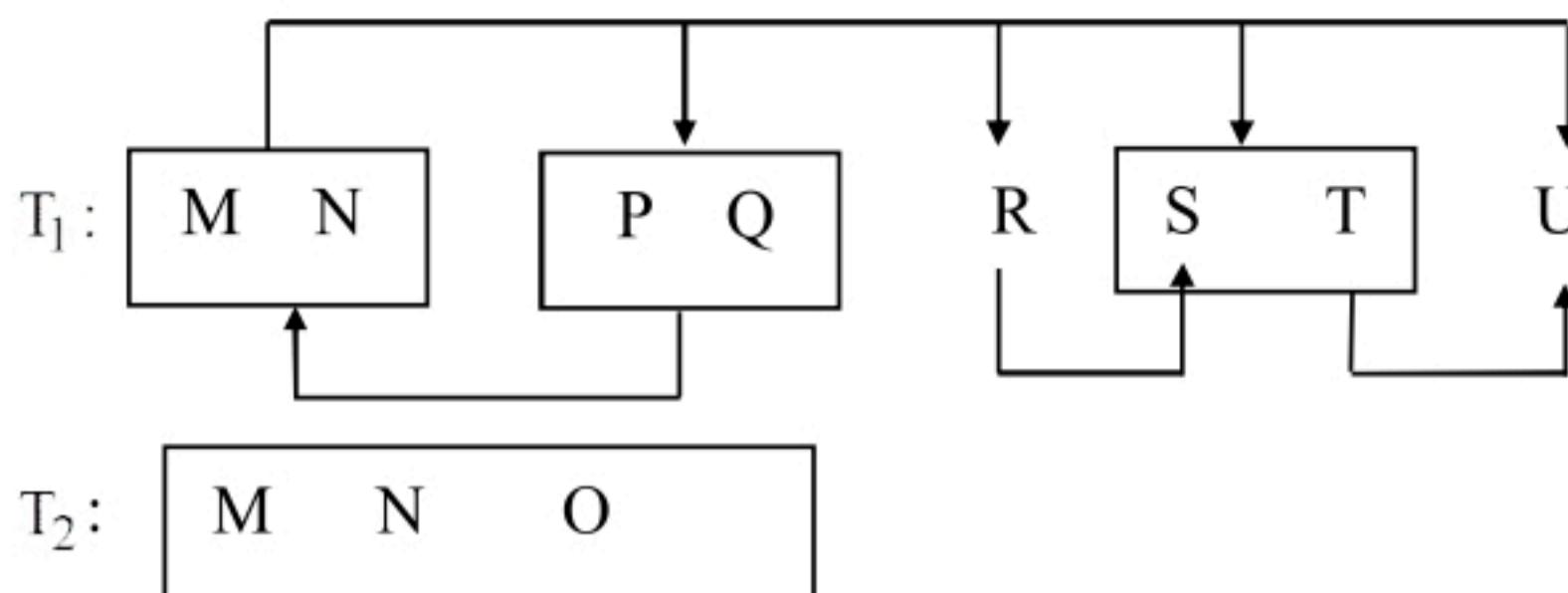
شکل ۸-۵. مثال ترکیبی برای فرم‌های نرمال

حل: این شکل در فرم نرمال 1NF نیست، زیرا صفت O به کلید وصل نیست،  
پس ابتدا آن را اصلاح می‌کنیم.



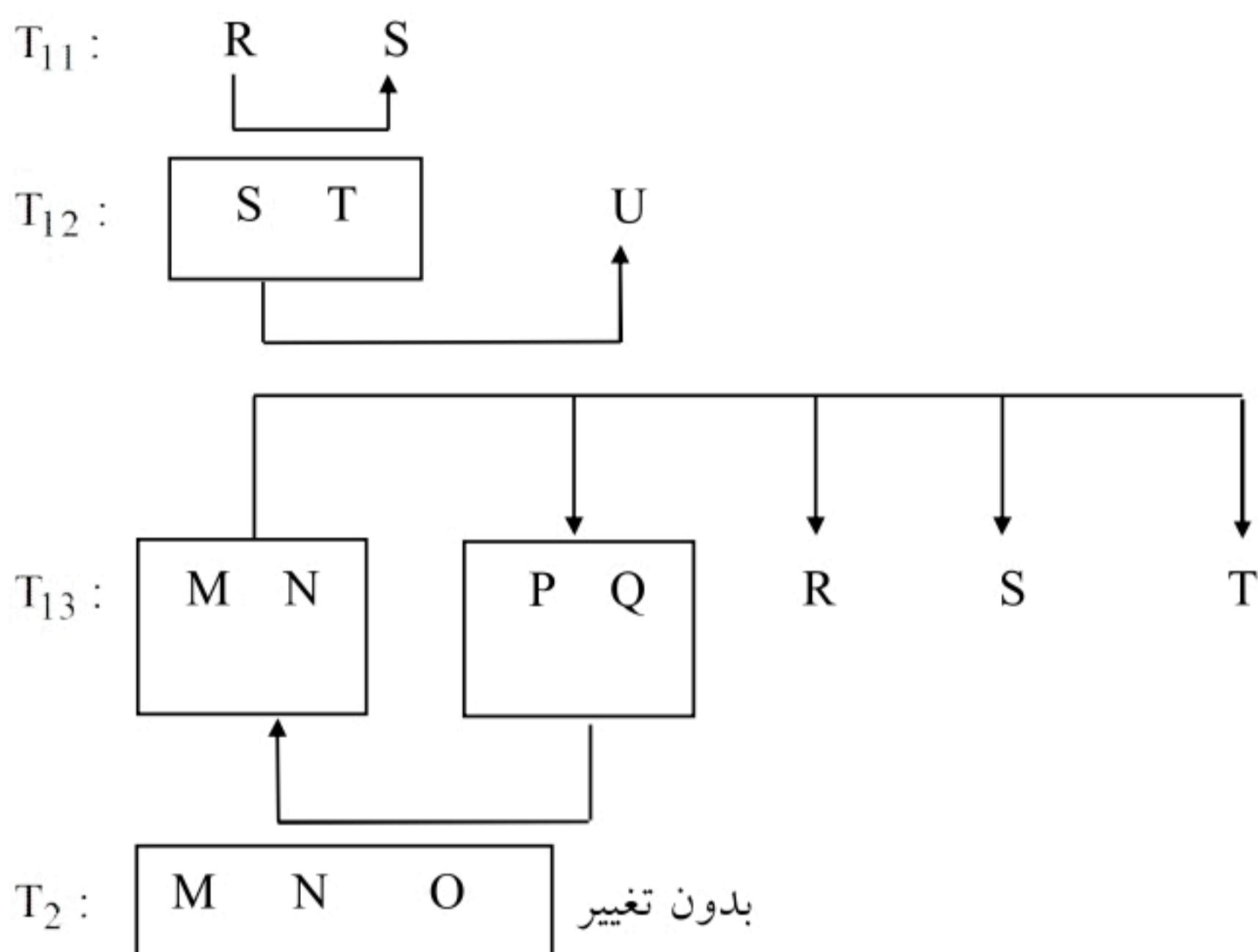
حالا این شکل معادل در فرم نرمال 1NF می‌باشد.

۱. تبدیل به 2NF



قواعد تجزیه رعایت و تمام وابستگی‌ها حفظ شده است.

### ۲. تبدیل به 3NF

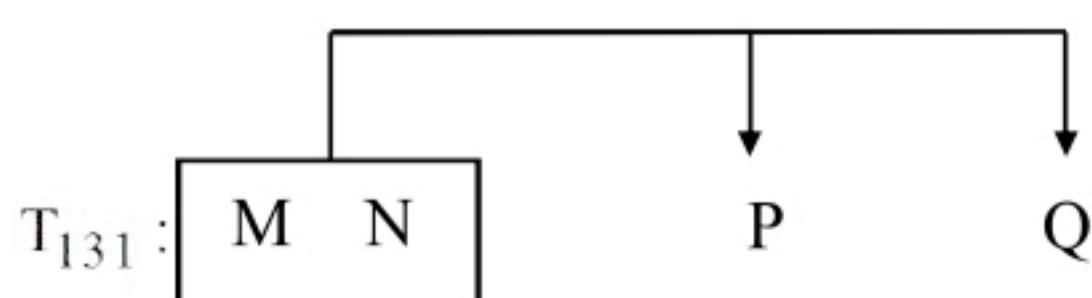


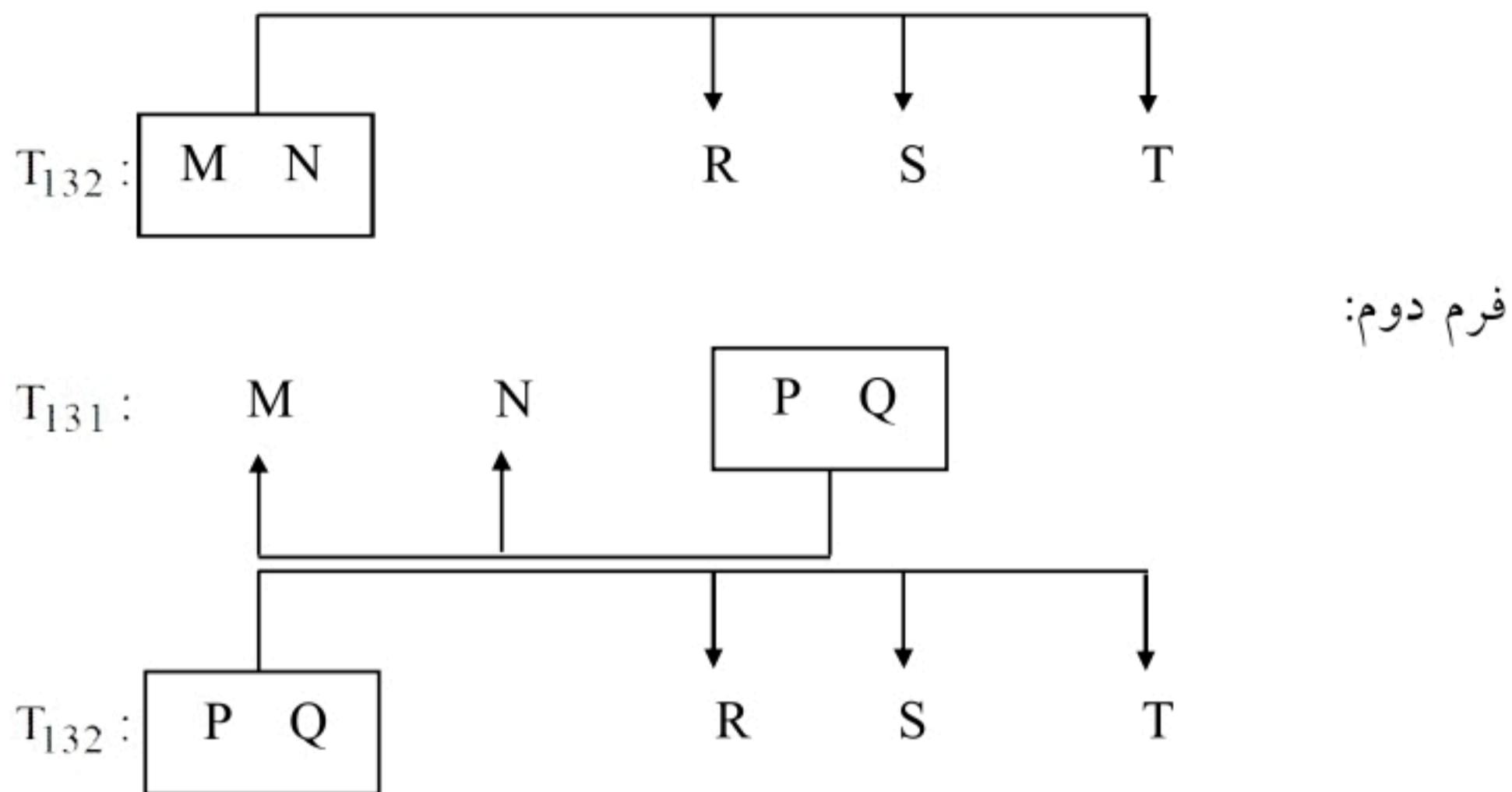
در این فاز نیز قواعد تجزیه رعایت و تمام وابستگی‌ها حفظ شده است.

### ۳. تبدیل به BCNF

تمام جداول حاصل، به جز  $T_{13}$  نرمال هستند. این جدول به صورت‌های زیر قابل تجزیه است:

فرم اول:





تصمیم‌گیری در مورد این دو فرم با طراح است. فرم اول بهتر است زیرا به تصمیم قبلی طراح در مورد کلید اصلی احترام می‌گذارد.

یادآوری: تجزیه T13 زمانی لازم است که در کاربرد مورد نظر، ستون‌های M, N, P, Q مستقل‌باشند، یعنی پرسش‌هایی که فقط به این ستون‌ها مربوط می‌شود مطرح شوند. در غیر این صورت نیازی به تجزیه جدول T13 نیست، زیرا کاربردی نخواهد داشت.

## ۵-۵ وابستگی چندمقداری و فرم نرمال چهارم

وابستگی چندمقداری، نوعی وابستگی بین دو مجموعه مستقل از صفت‌ها است که از وابستگی تابعی عامتر است. این حالت در مثال زیر مشاهده می‌شود.  
مثال:

جدول شکل ۹-۵ شامل نام دانشجو، اساتیدی که با آن‌ها درس دارد و وام‌هایی که دریافت کرده و تاریخ آن وام‌ها را در نظر بگیرید:

در این مثال، نام اساتید و وام‌های دریافت شده از یکدیگر مستقل هستند یعنی وابستگی تابعی ندارند.

اگر یک دانشجو چند استاد داشته باشد و چند وام دریافت کرده باشد، افزونگی داده ایجاد می‌شود.

Sname	Proaf	Loan	Date
علی	حسنی	ضروری	۱۳۹۳
علی	رضوی	ضروری	۱۳۹۴
علی	رضوی	ضروری	۱۳۹۳
علی	رضوی	مسکن	۱۳۹۴
علی	حسنی	مسکن	۱۳۹۴

شکل ۵-۹. مثال برای وابستگی چندمقداری و فرم نرمال چهارم (4NF)

در جدول فوق مشاهده می‌شود که نام اساتید و وام‌های «علی» تکرار شده‌اند (افزونگی). این در حالی است که جدول فوق تا سطح BCNF نرمال‌سازی شده‌است، زیرا هیچ‌گونه وابستگی تابعی به کلید اصلی، که شامل تمام صفت‌ها است، وجود ندارد.

این نوع وابستگی و فرم نرمال مربوط به آن به صورت‌های زیر تعریف می‌شود:

هرگاه دو ارتباط مستقل، بین مجموعه صفت‌های یک رابطه R مثل A:C و A:B وجود داشته باشد، وابستگی چندمقداری در رابطه R برقرار است که به دو صورت زیر نشان داده می‌شود:

$$A \rightarrow\rightarrow B, A \rightarrow\rightarrow C \text{ یا } A \rightarrow\rightarrow B | C$$

وابستگی چندمقداری به شکل‌های متفاوت در منابع مختلف تعریف شده‌است. یک تعریف دیگر هم در زیر آمده است:

در رابطه (R(A,B,C) داریم  $A \rightarrow\rightarrow B | C$ ، اگر هر مقدار A به طریقی به مقادیر B وابسته باشد که به C وابسته نباشد.

مثال: در جدول فوق وابستگی‌های چندمقداری زیر وجود دارند:  $sname \rightarrow\rightarrow prof$ ,  $sname \rightarrow\rightarrow loan$ ,  $sname \rightarrow\rightarrow date$  یا  $sname \rightarrow\rightarrow prof | loan | date$

تعریف:

یک وابستگی چندمقداری  $B \rightarrow\rightarrow A$  در رابطه R را بدیهی<sup>1</sup> گویند اگر یکی از دو شرط زیر برقرار باشد:

$$B \subset A .\ 1$$

$$A \cup B = R .\ 2$$

اگر در یک وابستگی چندمقداری، هیچ‌کدام از دو شرط فوق برقرار نباشد، به آن وابستگی چندمقداری غیربدیهی<sup>۱</sup> گویند. در جدول فوق هر دو وابستگی چندمقداری، غیربدیهی هستند.

برای محاسبه  $F^{opt}$  در وابستگی چندمقداری، قواعد مشخصی وجود دارد [H. F. Korth and A. Silberchatz 2006]

تعریف: رابطه  $R$  در  $^2 4NF$  است در صورتی که اگر وابستگی چندمقداری غیربدیهی  $\rightarrow BA$  وجود داشته باشد، آنگاه  $A$  ابرکلید  $R$  باشد.

جدولی که دارای وابستگی چندمقداری هستند را می‌توان به صورت زیر تجزیه کرد و افزونگی را از بین برد.

جدول اول شامل مجموعه صفت‌های  $A$  و  $B$

جدول دوم شامل مجموعه صفت‌های  $A$  و  $C$

مثال:

جدول فوق به دو جدول زیر تجزیه می‌شود و افزونگی از بین می‌رود.

Sname	loan	date
علی	مسکن	۱۳۹۳
علی	مسکن	۱۳۹۴
علی	ضروری	۱۳۹۴

Sname	prof
علی	حسنی
علی	رضوی

## ۶-۵ وابستگی پیوندی و فرم نرمال پنجم<sup>۲</sup>

با توجه به شرایط صحت تجزیه، در مواردی نمی‌توان جدولی را به‌طور صحیح به دو جدول تجزیه کرد، ولی تجزیه آن به سه جدول یا بیشتر امکان‌پذیر است. چنین جداولی دارای وابستگی پیوندی هستند.

- 
1. Non-Trivial
  2. Fourth Normal Form
  3. Fifth Normal Form

تعریف:

اگر  $R$  یک رابطه و ستون‌های هریک از رابطه‌های  $A, B, \dots, P$  زیرمجموعه ستون‌های  $R$  باشند، آنگاه  $R$  دارای وابستگی پیوندی روی رابطه‌های  $A, B, \dots, P$  است، اگر و تنها اگر داشته باشیم:

$$R = A \bowtie B \bowtie \dots \bowtie P$$

تعریف:

جدول  $R$  در 5NF است اگر و تنها اگر فقط به کلیدهای کاندیدش وابستگی پیوندی داشته باشد.

شکل‌های نرمال دیگری چون GNF و DKNF و RUNF و ... وجود دارند [روحانی، ۱۳۹۲] ولی چندان مطرح نیستند و کاربرد ندارند. باید تأکید کرد که با افزایش شماره و پیچیدگی، فرم‌های نرمال کوچکتر می‌شوند (1NF بزرگ‌ترین مجموعه جداول نرمال است)، فرمول زیر:

$$\dots 5NF \subseteq 4NF \subseteq BCNF \subseteq 3NF \subseteq 2NF \subseteq 1NF$$

مثال: جدول ABC در شکل ۱۱-۵ را که در آن هر سه ستون با یکدیگر کلید کاندید را تشکیل می‌دهند، می‌توان به سه جدول A و B و C تجزیه کرد، ولی تجزیه آن به طور صحیح به دو جدول امکان ندارد.

**ABC**

A	B	C
a1	b1	c2
a1	b2	c1
a2	b1	c1
a1	b1	c1

A	B
a1	b1
a1	b2
a2	b1

B	C
b1	c2
b2	c1
b1	c1

A	C
a1	c2
a1	c1
a2	c1

join

join

جدول اصلی  
ABC

سطر اضافی

A	B	C
a1	b1	c2
a1	b1	c1
a2	b1	c2
a2	b1	c1
a1	b2	c1

شکل ۱۰-۵. مثال برای وابستگی پیوندی و فرم نرمال پنجم

## خلاصه فصل پنجم

این فصل نیز به مدل رابطه‌ای اختصاص دارد. رابطه‌ها اگر درست طراحی نشوند، می‌توانند حافظه و زمان اجرا را هدر دهند. اصول طراحی رابطه‌ها نرمال‌سازی نام دارد. در این فصل ابتدا مشکلات جداول غیرنرمال آشکار شد و سپس به مبانی نظری نرمال‌سازی پرداخته شد و آنگاه چگونگی اعمال نرمال‌سازی به صورتی مبسوط بیان شد. مفهوم وابستگی، بهویژه وابستگی تابعی، که مبنای نرمال‌سازی است ارائه شده و ابزاری چون کلید و نمودار وابستگی تابعی به کار گرفته شده است. بر اساس انواع وابستگی، فرم‌های نرمال تعریف شد و چگونگی شکستن (تجزیه) جداول غیرنرمال و تبدیل آن‌ها به جداول نرمال همراه با الگوریتم‌ها و روش‌های شناخت شده بیان شده است. یک جدول را می‌توان به شکل‌های متفاوتی تجزیه کرد. کدام تجزیه درست و کدام نادرست است؟ این مطلب در قالب قواعد صحت تجزیه جدول ارائه شده تا از اشتباهات معمول جلوگیری شود. در اینجا هم مبنای کار همان وابستگی است، یعنی مجموعه جداول به دست آمده از تجزیه یک جدول باید همان وابستگی‌های قبلی را حفظ کنند و نیز قابل پیوند به جدول اصلی باشند. در پایان این فصل، وابستگی‌های چندمقداری و پیوندی که نسبت به وابستگی تابعی کمیاب‌تر هستند، ارائه شده‌اند.

## خودآزمایی تشریحی فصل پنجم

۱. در پایگاهداده زیر ابتدا کلیدهای کاندید را بیابید و سپس آن را به‌طور کامل

نرمال‌سازی کنید:

$$R = \{A, B, C, D, E, F, G\}$$

$$F = \{AF \rightarrow BE, FC \rightarrow DE, F \rightarrow CD, D \rightarrow E, C \rightarrow A\}$$

حل: با توجه به پیچیدگی  $F$ ، ابتدا آن را ساده می‌کنیم و سپس به پیدا کردن  $F^+$  می‌پردازیم. سمت راست وابستگی‌ها را به یک صفت تبدیل می‌کنیم.

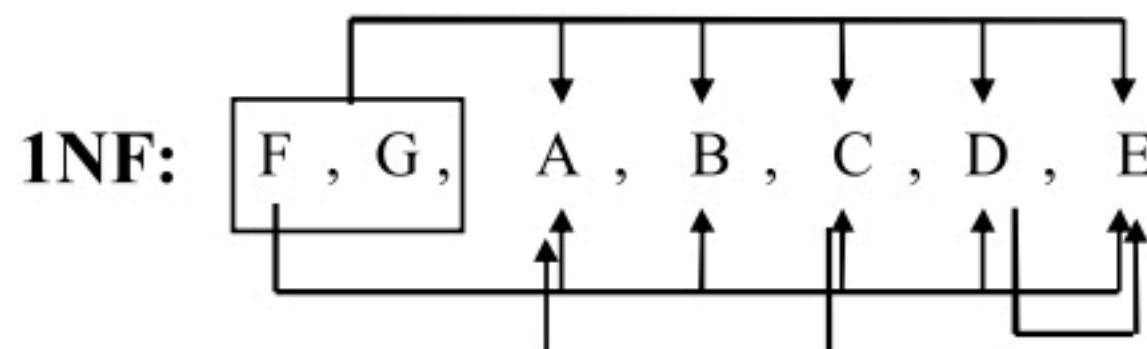
$$AF \rightarrow B \quad AF \rightarrow E \quad FC \rightarrow D \quad FC \rightarrow E$$

$$F \rightarrow C \quad F \rightarrow D \quad D \rightarrow E \quad C \rightarrow A$$

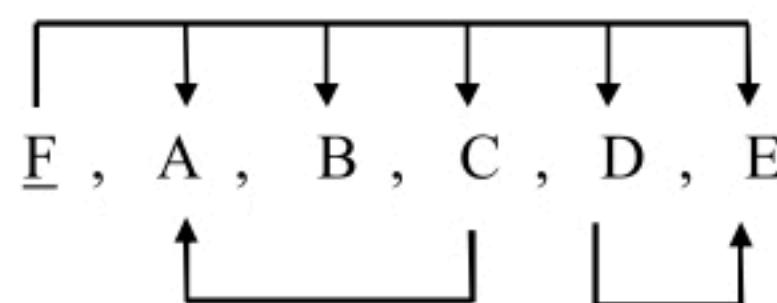
می‌توان نتیجه گرفت که

$$F \rightarrow A \quad \text{و} \quad F \rightarrow B$$

صفت F همه صفت‌های دیگر، به جز G، را می‌دهد. پس (F,G) کلید کاندید است. این کلید کاندید منحصر به فرد است، زیرا هیچ صفتی F و G را نمی‌دهد (یعنی در هر کلید کاندید، این دو صفت لازم هستند).



**2NF:** (F, G)



**3NF:** (F, D)

(C,A)

(D,E)

(F,B,C,D)

نیازی به BCNF نیست، زیرا بیش از یک کلید کاندید نداریم.

۲. پایگاهداده زیر را در نظر بگیرید:

$$R = \{X, Y, Z, S, T, U, W\}$$

$$F = \{S \rightarrow X, T \rightarrow Y, X \rightarrow Y, XY \rightarrow TUZ\}$$

الف) کلیدهای کاندید را بیابید.

ب) به طور کامل نرم‌السازی کنید.

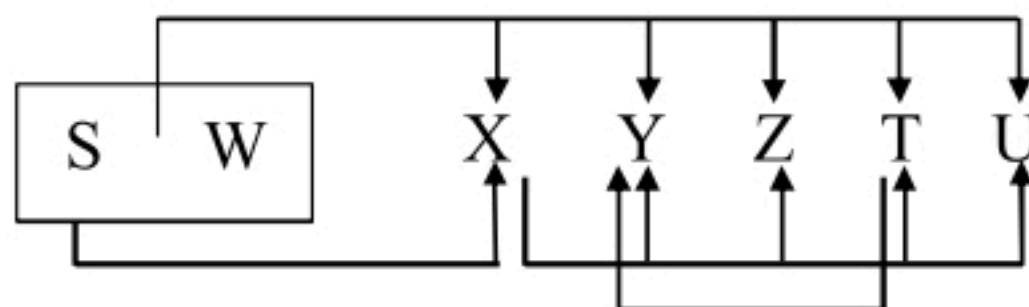
حل:

$$S \rightarrow X, X \rightarrow Y \Rightarrow S \rightarrow XY \quad S \rightarrow XY, XY \rightarrow TUZ \Rightarrow S \rightarrow T, S \rightarrow U, S \rightarrow Z$$

پس (S,W) تنها کلید کاندید است.

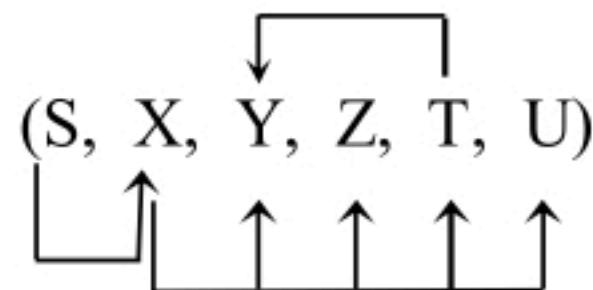
نمودار وابستگی تابعی در زیر آمده است:

1NF :



فرم‌های نرمال:

2NF : (S, W)



3NF :

- (T, Y)
- (X, Z, T, U)
- (S, X)

نیازی به BCNF نیست، زیرا بیش از یک کلید کاندید نداریم.

باید تأکید کرد که مجموعه وابستگی پوششی بهینه نقش مهمی در نرمال‌سازی دارد. اگر از بعضی وابستگی‌های آن صرف‌نظر کنیم، به نتیجه غلطی می‌رسیم.

### تمرین‌های تشریحی فصل پنجم

۱. در یافتن مجموعه پوششی وابستگی چه نیازی به قواعد ۴ و ۵ و ۶ داریم؟
۲. از روی نمودار وابستگی شکل ۲-۵ وابستگی‌های مربوطه را استخراج کنید.
۳. فرم‌های نرمال 2NF و 3NF را در هم ادغام کنید و الگوریتمی برای گذار از 1NF به 3NF پیشنهاد کنید.
۴. در الگوریتم گذار از 1NF به 2NF، شماره ۳ چه اهمیتی دارد؟ مثالی بزنید. همچنین برای الگوریتم گذار از 2NF به 3NF.
۵. بانک اطلاعات ثبت‌نام دانشگاه را بدون رعایت مراحل ۱ و ۲ در گذار به نرمال‌سازی و نتیجه را بررسی کنید.
۶. اگر بانکی در 3NF باشد ولی در BCNF نباشد، چه اشکالی دارد؟
۷. برای هریک از موارد ارائه شده در بخش BCNF مثال دیگری پیشنهاد کنید.
۸. بانک اطلاعات زیر را نرمال‌سازی کنید و درستی کار خود را بسنجید.

$$R = (P, Q, S, T, U, V)$$

$$F = \{P \rightarrow QV, S \rightarrow T, T \rightarrow PU, Q \rightarrow V, Q \rightarrow U\}$$

۹. در بانک اطلاعات زیر همه کلیدهای کاندید را بیابید. نمودارهای وابستگی تابعی را رسم کنید:

$$R = (A, B, C, D, E, F, G, H)$$

$$F = \{ABC \rightarrow DF, DF \rightarrow G, F \rightarrow B, D \rightarrow B, G \rightarrow H, B \rightarrow C\}$$

۱۰. وابستگی  $C \rightarrow B$  را از تمرین ۹ حذف و آن را حل کنید.

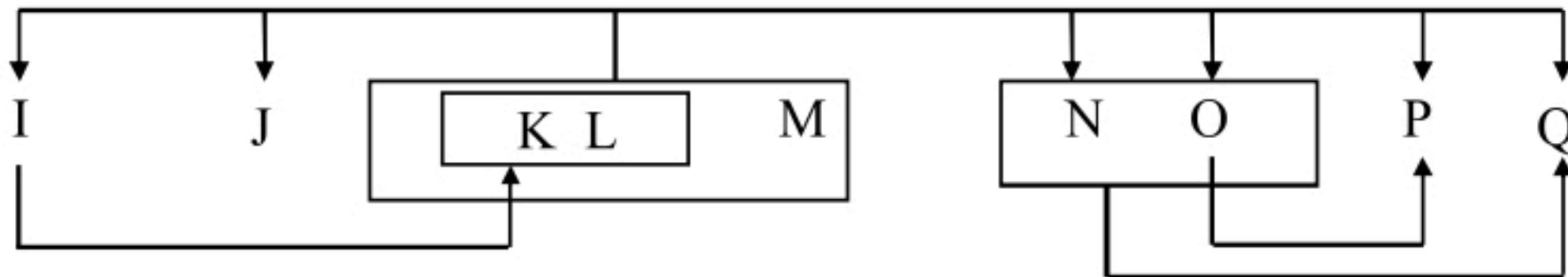
۱۱. در بانک اطلاعات زیر همه کلیدهای کاندید را بیابید و نمودار وابستگی تابعی را رسم کنید:

$$R = (I, J, K, L, M, N, O, P)$$

$$F = \{IJK \rightarrow R, IJ \rightarrow O, J \rightarrow P, N \rightarrow I\}$$

۱۲. در مسئله ۱۱ وابستگی‌های  $I \rightarrow O$  و  $J \rightarrow O$  را حذف و وابستگی  $KL \rightarrow O$  را اضافه و مسئله را حل کنید.

۱۳. بانک اطلاعات زیر را به‌طور کامل نرمال‌سازی کنید.



۱۴. رابطه زیر و وابستگی‌های آن را در نظر بگیرید:

$$R = (A, B, C, D, E, F, G)$$

$$F = \{AF \rightarrow BE, FC \rightarrow DE, F \rightarrow CD, D \rightarrow E, C \rightarrow A\}$$

الف) کلیدهای کاندید را بیابید.

ب) این رابطه را به‌طور کامل نرمال‌سازی کنید.

ج) نشان دهید که در نرمال‌سازی فوق، اطلاعات از دست نرفته است.

رابطه  $P(X, Y, Z, T, S)$  با وابستگی‌های زیر موجود است:

$$X \rightarrow YZ, ZT \rightarrow S, Y \rightarrow T, S \rightarrow X$$

این رابطه را تا سطح BCNF قدم به قدم نرمال‌سازی کنید.