

نرمال سازی (Normalization)

نرمال سازی-مقدمه

- هنگام طراحی یک پایگاه داده مواردی مطرح می شود:

- ✓ با توجه به داده های عملیاتی و ارتباط بین موجودیت ها، چه تعداد جدول باید طراحی شود؟

- ✓ در هر جدول چه فیلدهایی تعریف شود؟

- ✓ رابطه جدول ها با هم چگونه باشد؟

- ✓ میزان افزونگی داده ها تا چه حد قابل قبول است؟

- برای پاسخ به این سوالات نیازمند بهره گیری از قواعد و قوانین نرمال سازی هستیم.

نرمال سازی-مقدمه

- پایگاه داده تهیه کننده قطعه:

این پایگاه داده را بر اساس ۳ جدول مقابل طراحی شده است.

$S (S \#, Sname, Status, City)$

$P (P \#, Pname, Color, Weight, City)$

$SP (S \#, P \#, QTY)$

Redesign



$S' (S \#, Sname, City)$

$P (P \#, Pname, Color, Weight, City)$

$SP' (S \#, P \#, QTY, Status)$

حال تغییری در طراحی این پایگاه داده می دهیم و نتیجه این تغییر را بررسی می کنیم:

جدول S و SP را در نظر بگیرید. از جدول S، فیلد City را حذف کرده و به جدول SP اضافه می کنیم.

نرمال سازی-مقدمه

SP

| S# | P# | QTY | Status |
|----|----|-----|--------|
| S1 | P1 | 300 | 20 |
| S1 | P2 | 200 | 20 |
| S1 | P3 | 400 | 20 |
| S1 | P4 | 200 | 20 |
| S1 | P5 | 100 | 20 |
| S1 | P6 | 100 | 20 |
| S2 | P1 | 300 | 10 |
| S2 | P2 | 400 | 10 |

- افزودن اطلاعات مشهود است:

مثلا مقادیر Status بی دلیل به ازای S# های مشخص تکرار شده است.

ضرورت نرمال سازی

- آیا می توان تمام اطلاعات سه جدول S، P و SP را در یک جدول ریخت (Flat File یا One Table)؟
بله شدنی است و از دید کاربران و طراحان ساده بوده و برای دستیابی به اطلاعات نیازی به انجام پیوند بین جداول نیست.

- معایب One Table یا ضرورت نرمال سازی:

۱- افزونگی داده ها (Data Redundancy)

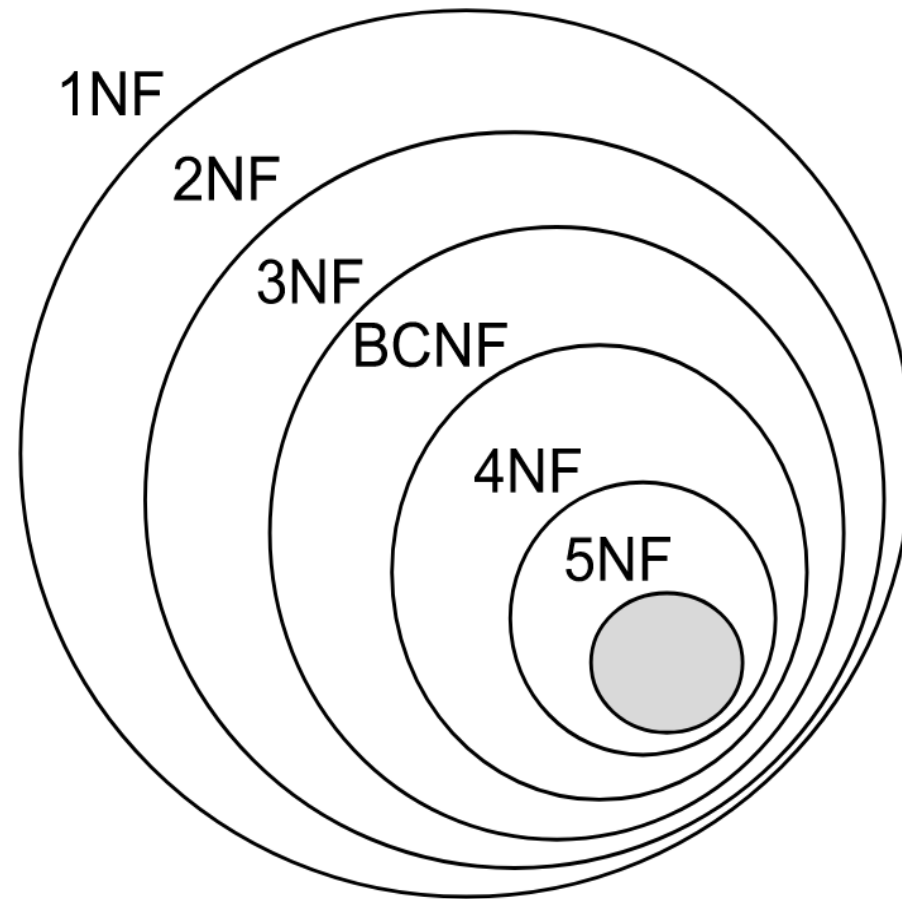
۲- آنومالی یا ناهنجاری یا بی نظمی (Anomaly)

۳- مقادیر تھی (Null Value)

S,P,SP

[illegible]

سطوح نرمال سازی



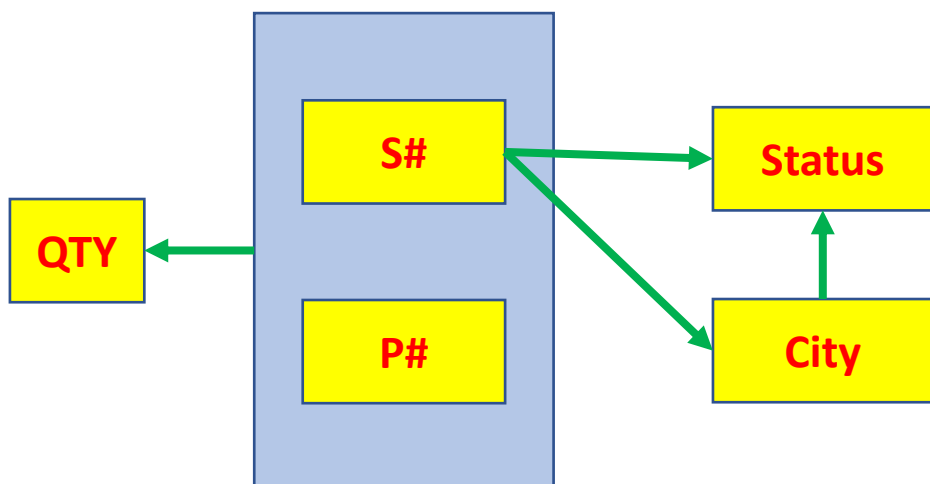
فرم نرمال 1NF

- رابطه R به فرم نرمال 1NF است اگر و فقط اگر تمام صفات آن روی میدان های تجزیه ناپذیر (اتومیک) تعریف شده باشد.
- مثال: وجود فیلد تاریخ که از سه فیلد کوچکتر (سال، ماه، روز) تشکیل شده است، فرم نرمال 1NF را از بین می برد.
- الگوریتم تبدیل رابطه غیر نرمال به 1NF:
 - ۱- صفات (فیلد های) مرکب را جدا کنید (ستون های جدید به جدول اضافه می شود).
 - ۲- صفات چند مقداری را گسترش دهید (اگر صفتی دو مقدار دارد آنگاه یک سطر تبدیل به دو سطر می شود).

فرم نرمال 2NF

- فرض کنید به جای دو جدول S و SP یک جدول به نام F به صورت زیر تعریف شود:

First (S #, Status, City, P #, QTY)



- کلید اصلی رابطه (S#, P#)
- وابستگی مقابل برقرار است:

$City \rightarrow Status$

❖ معایب طراحی:

-افزونگی اطلاعات

وجود افزونگی باعث آنومالی (ناهنجاری) می شود.

فرم نرمال 2NF - ادامه

- انواع آنومالی که به علت افزونگی اطلاعات ایجاد می شود:

۱- آنومالی درج:

افزودن اطلاع "S7 در شهر C4 ساکن است"، تا زمانی که مشخص نشود چه قطعه ای تولید کرده است، امکان پذیر نمی باشد (P# جز کلید اصلی است و نمی تواند Null باشد).

۲- آنومالی حذف:

اگر اطلاع "S3 از P2 به تعداد ۲۰۰ عدد تهیه کرده است" را حذف کنیم، اطلاع "S3 ساکن شهر C3 است" نیز ناخواسته حذف خواهد شد.

۳- آنومالی به روز رسانی:

مشکل به روز رسانی منتشر شونده به دلیل دفعات تکرار یک مقدار رخ می دهد و اگر انجام نشود پایگاه داده در وضعیت ناسازگار قرار می گیرد.

First

| S# | Status | City | P# | QTY |
|----|--------|------|----|-----|
| S1 | 20 | C2 | P1 | 300 |
| S1 | 20 | C2 | P2 | 200 |
| S1 | 20 | C2 | P3 | 400 |
| S2 | 10 | C3 | P1 | 100 |
| S2 | 10 | C3 | P2 | 200 |
| S3 | 10 | C3 | P2 | 200 |
| S4 | 20 | C2 | P2 | 200 |
| S4 | 20 | C2 | P4 | 300 |

فرم نرمال 2NF - ادامه

- دلیل آنومالی های ذکر شده به خاطر این است که فیلد های Status و City با کلید اصلی وابستگی تابعی کامل ندارند.

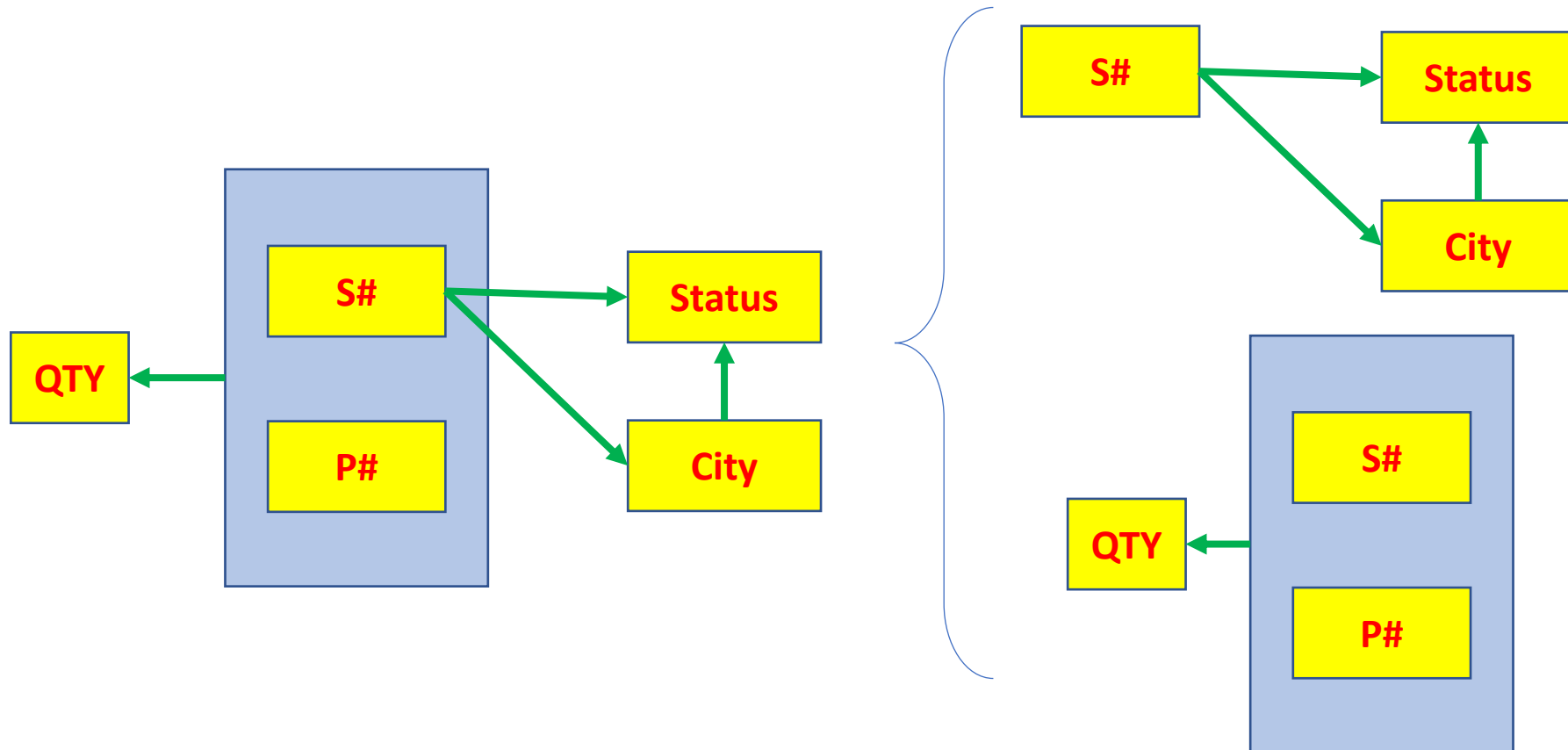
• تعریف:

رابطه R در صورت دوم نرمال است اگر و فقط اگر 1NF باشد و هر صفت غیر کلید با کلید اصلی وابستگی تابعی کامل داشته باشد (هر صفت غیر کلید با کلید اصلی به طور کاهش ناپذیر وابسته باشد).

- صفت غیر کلید صفتی است که خود کلید نباشد و جزء تشکیل دهنده کلید هم نباشد.
- صفتی که جزئی از کلید باشد را صفت عمده (Prime Attribute) می نامیم.

فرم نرمال 2NF - ادامه

First ($S \#, Status, City, P \#, QTY$)
Second ($S \#, Status, City$)
 SP ($S \#, P \#, QTY$)



فرم نرمال 3NF

SP

| S# | P# | QTY |
|----|----|-----|
| S1 | P1 | 300 |
| S1 | P2 | 200 |
| S1 | P3 | 400 |
| S1 | P4 | 200 |
| S1 | P5 | 100 |
| S1 | P6 | 100 |
| S2 | P1 | 300 |
| S2 | P2 | 200 |
| S3 | P2 | 200 |
| S4 | P2 | 200 |
| S4 | P4 | 300 |
| S4 | P5 | 400 |

- مشکلات ناشی از وابستگی غیر کامل به کلید با 2NF نمودن رابطه (عمل پرتو (تجزیه جدول)) حل شد.

- ولی جدول Second همچنان موجب برخی آنومالی ها می شود:

- مثال:

Second

| S# | Status | City |
|----|--------|------|
| S1 | 20 | C2 |
| S2 | 10 | C3 |
| S3 | 10 | C3 |
| S4 | 20 | C2 |
| S5 | 30 | C1 |

نمی توان اطلاع "یک شهر خاص دارای یک مقدار وضعیت مشخص است" را تا زمانی که ندانیم چه تهیه کننده ای در شهر ساکن است، در جدول Second درج کنیم.

همچنین اگر تاپلی از جدول Second حذف کنیم، نه تنها تهیه کننده ای را حذف کرده ایم بلکه اطلاع در مورد یک شهر و وضعیت آن را نیز حذف نموده ایم.

فرم نرمال 3NF - ادامه

- دلیل آنومالی ذکر شده در اسلاید قبل، وابستگی با واسطه است.
- تعریف:

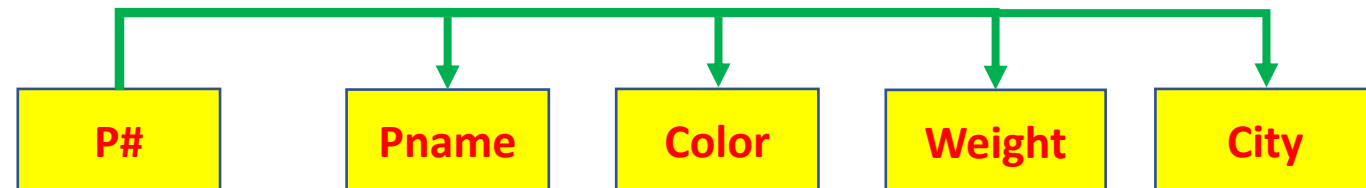
رابطه R در سطح نرمال 3NF است اگر و فقط اگر صفات غیر کلید:
۱- متقابلاً به یکدیگر ناوابسته باشند.

۲- با کلید اصلی رابطه R وابستگی تابعی کامل داشته باشند.

- مثال:

رابطه P در سطح نرمال سوم است. چون همه صفات با کلید اصلی رابطه (P#) وابستگی تابعی کامل دارند و همچنین همه صفات غیر کلید به یکدیگر ناوابسته هستند.

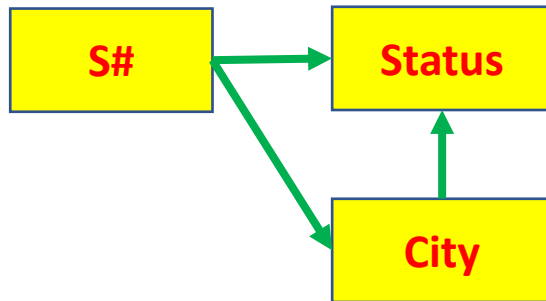
P (P#, Pname, Color, Weight, City)



فرم نرمال 3NF - ادامه

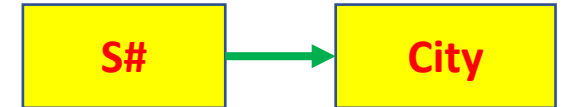
- مثال: جدول Second در فرم نرمال ۲ است ولی 3NF نیست.
- زیرا فیلد غیر کلید Status وابسته به فیلد غیر کلید City می باشد.

Second

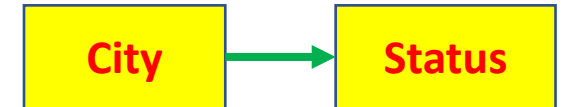


Second ($S \#, Status, City$)

SC ($\underline{S \#}, City$)



CS ($City, \underline{Status}$)



- تعریف:

رابطه R در فرم سوم نرمال (3NF) است، اگر و فقط اگر 2NF بوده و هر صفت غیر کلید بطور مستقیم (بی واسطه) با کلید اصلی وابستگی داشته باشد (حذف وابستگی های با واسطه).

فرم نرمال BCNF

- تعریف:

رابطه R در سطح Boyce-Codd Normal Form (BCNF) است اگر و فقط اگر هر دترمینان، کلید کاندید باشد. به عبارتی دیگر، جدولی در BCNF است که ستون های آن فقط به کلید های کاندیدش وابستگی تابعی داشته باشند.

دترمینان: هر صفتی که صفتی دیگر با آن وابستگی تابعی کامل داشته باشد، دترمینان نامیده می شود. یعنی اگر داشته باشیم $A \rightarrow B$ و این وابستگی کامل باشد، صفت A را دترمینان می نامیم.

- نکته: در تعریف BCNF تاکید روی کلید کاندید است نه لزوما کلید اصلی

نکته: رابطه ای که 3NF نباشد، BCNF نیز نخواهد بود.

هر رابطه BCNF حتما 3NF است ولی هر رابطه 3NF لزوما BCNF نیست.

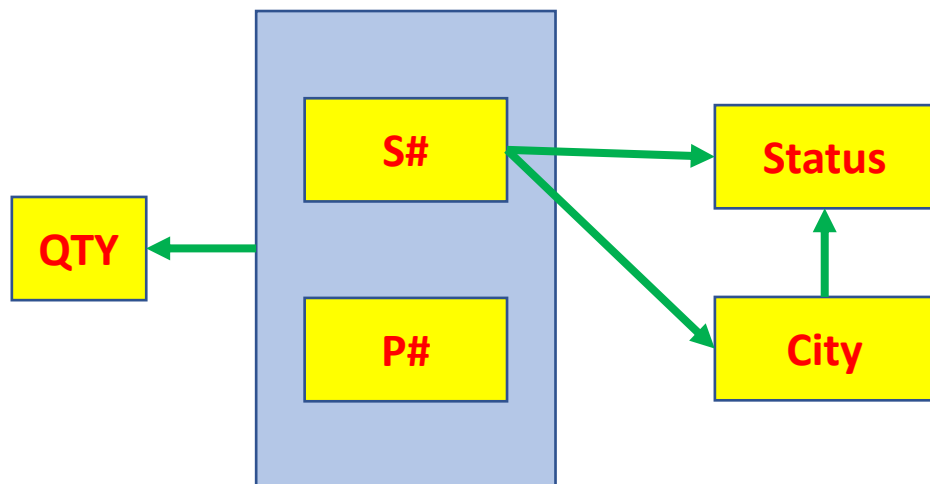
فرم نرمال BCNF-ادامه

• مثال:

رابطه $SP(S\#,P\#,QTY)$ در سطح نرمال BCNF است. زیرا با توجه به اینکه کلید کاندید و اصلی رابطه $(S\#,P\#)$ است و QTY فقط با آن وابستگی تابعی دارد، بنابراین در سطح BCNF می باشد.

• مثال:

رابطه $First(S\#,Status,City,P\#,QTY)$ در سطح BCNF نیست چرا که دارای سه دترمینان $S\#$ و $City$ و $(S\#,P\#)$ است که دو تای آنها کلید کاندید نیستند ($S\#$ و $City$):



مثال:

- رابطه R بصورت زیر و با FD هایی مشخص شده است:

$R (X, Y, Z, S, T, U, W)$

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

الف) کلید های کاندید را بیابید.

ب) رابطه را نرمال سازی نمائید.

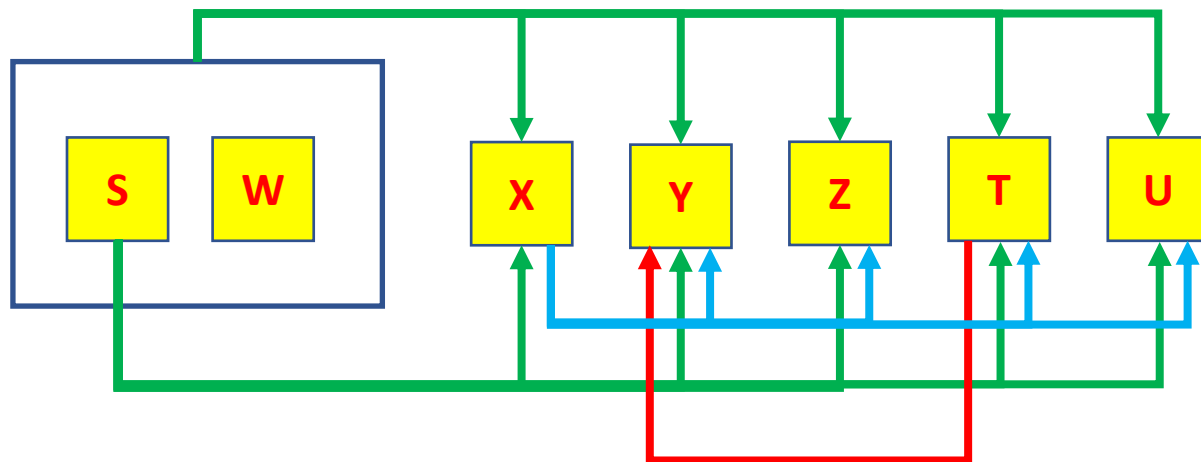
$$S \rightarrow X, X \rightarrow Y \Rightarrow S \rightarrow Y \Rightarrow S \rightarrow XY$$

$$S \rightarrow XY, XY \rightarrow TUZ \Rightarrow S \rightarrow TUZ \Rightarrow S \rightarrow T, S \rightarrow U, S \rightarrow Z$$

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

$$F_{OPT} = \{S \rightarrow X, S \rightarrow Y, S \rightarrow T, S \rightarrow U, S \rightarrow Z, T \rightarrow Y, X \rightarrow Y, X \rightarrow T, X \rightarrow U, X \rightarrow Z\}$$

S تمام صفات دیگر به جز W را می دهد، بنابراین (S,W) کلید کاندید است.



رابط با توجه به نمودار وابستگی و مجموعه F در سطح 1NF است.

$$2NF \left\{ \begin{array}{l} (\underline{S}, W) \\ (\underline{S}, X, Y, Z, T, U) \end{array} \right. :2NF$$

$$3NF \left\{ \begin{array}{l} (\underline{S}, W) \\ (\underline{S}, X) \\ (\underline{X}, Z, T, U) \\ (\underline{T}, Y) \end{array} \right. :3NF$$

نیازی به BCNF نیست زیرا در هر رابطه بیش از یک کلید کاندید نداریم و روابط 3NF فوق در سطح BCNF هستند.

نکته:

- برای نرمال سازی حتما باید مجموعه وابستگی های پوششی بهینه را بدست آوریم.
- رابطه 3NF که تنها یک کلید کاندید دارد، BCNF است.
- در صورت نبود صفت مشترک بین کلید های کاندید، اگر رابطه 3NF باشد، BCNF هم هست.