

# اصول طراحی پایگاه داده

By Dr. Taghinezhad

University of Tabriz, Fall 2024

Mail:

[a0taghinezhad@gmail.com](mailto:a0taghinezhad@gmail.com)

# Normalization

نرمال سازی

# مباحث

- معرفی نرمال سازی
- هدف از نرمال سازی
- سطوح مختلف نرمال سازی
- معایب نرمال سازی

# جداول آنرمال

## ■ نرمال سازی

- روشی برای طراحی جداول پایگاه داده و داده ها
- به طریقی که باعث کاهش افزونگی داده
- رفع مشکلات ساختاری و آنومالی

# جداول آنرمال

■ هدف از نرمال سازی

- حذف افزونگی داده
- باقی نگاه داشتن وابستگی بین داده های مرتبط است.
- به این طریق اندازه پایگاه داده را کاهش داده و ذخیره منطقی داده را تضمین می کند.

# جداول آنرمال

## ■ فرآیند نرمال سازی

- شامل ایجاد جداول
- برقراری ارتباط بین آنها طبق قواعد معین است
- روی وابستگی های ستون های جدول تمرکز دارد.
- این فرآیند اغلب باعث ایجاد جداول بیشتر می شود
- باوجودیکه اثر تکرار داده درون پایگاه داده را دارد باعث افزونگی غیر ضروری داده نمی شود.

# مثال. جدول زیر که اطلاعات مربوط به خرید مشتریان را دارد در نظر بگیرید:

همانطور که مشاهده می شود با هر فروش داده ها در جدول تکرار می شوند.

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
12351	Aug 14 2002	AQX88916	55	1317.25	Dave Williams	6794-1674	Diane	Adams	364 East Road	150

## مثال. جدول زیر که اطلاعات مربوط به خرید مشتریان را دارد در نظر بگیرید:

این افزونگی مشکلات زیر را می تواند ایجاد کند:

### ■ هدر رفتن فضای ذخیره سازی.

- با وجودیکه امروزه دیسک های چندصد گیگا بایتی وجود دارد چندین بار ذخیره یک داده غیر ضروری است.

### ■ آنومالی در بهنگام سازی

- اگر داده یک مشتری، مثلا آدرس، تغییر کند باید در همه جاهائی که ذخیره شده است این تغییر اعمال شود درغیراینصورت جامعیت نقص می شود.

### ■ آنومالی در حذف.

- اگر این جدول به منظور نگهداری مشخصات مشتریان باشد، اگر مشتری خریدش را پس بدهد و سطر مربوط به آن حذف شود کلیه اطلاعات مشتری هم حذف می شود.

### ■ آنومالی در درج.

- به همین صورت نمی توانیم مشخصات مشتری جدید را درج کنیم مگر اینکه کالائی خریده باشد.

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
12351	Aug 14 2002	AQX88916	55	1317.25	Dave Williams	6794-1674	Diane	Adams	364 East Road	150



- راهکار:
- جدا کردن داده های جدول زیر به جداول جداگانه افزونگی را کاهش می دهد
- مواجهه با آنومالی های فوق را ساده تر می کند.
- این فرآیند را نرمالسازی می نامند.

مثال. جدول زیر که اطلاعات مربوط به خرید مشتریان را دارد در نظر بگیرید:

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
12351	Aug 14 2002	AQX88916	55	1317.25	Dave Williams	6794-1674	Diane	Adams	364 East Road	150

- تئوری پایگاه داده درجه نرمالسازی جدول را با اصطلاح فرم های نرمال (normal form) شرح می دهد. فرم های نرمال (یا بطور خلاصه NF) معیاری برای تعیین درجه نرمال جدول در اختیار می گذارد.
- فرم های نرمال جداگانه روی هر جدول می توانند بکار بروند. پایگاه داده زمانی در فرم نرمال  $n$  خواهد بود که کل جداول آن در فرم نرمال  $n$  باشند.
- فرم های نرمال عبارتند از:

- - First Normal Form (1NF)
  - Second Normal Form (2NF)
  - Third Normal Form (3NF)
  - Forth Normal Form (4NF)
  - Boyce/Codd Normal Form (BCNF)
  - Fifth Normal Form (5NF)
  - Domain/Key Normal Form (DKNF)

# جداول آنرمال

- جداول آنرمال به جداولی اطلاق میشود که در برخورد هر سطر با هر ستون آن به جای یک مقدار اتمی و تجزیه ناپذیر، مجموعه ای از مقادیر وجود دارد (مانند Telephones)

S#	Name	Telephones
7801	آرش	0311-6262778 0913-311-5234
7902	عسل	021-2956677 0912-314-4532

# جداول آنرمال

■ مهمترین عیب یک جدول آنرمال این است که برای هر یک از عملیات درج، حذف و اضافه به دو دسته عملیات درج تاپل و درج گروه اطلاعات مجموعه (Telephones) احتیاج است.

S#	Name	Telephones
7801	آرش	0311-6262778 0913-311-5234
7902	عسل	021-2956677 0912-314-4532

# جدول نرمال ۱

- یک جدول نرمال ۱ است اگر در برخورد هر سطر با هر ستون به یک مقدار تجزیه ناپذیر برسیم
- برای آنکه جدول آنرمال Student را به نرمال ۱ تبدیل کنیم، لازم است مقادیر ویژگیهای St# و Name را به ازاء هر شماره تلفن تکرار کنیم

S#	Name	Telephones
7801	آرش	0311-6262778
7801	آرش	0913-311-5234
7902	عسل	021-2956677
7902	عسل	0912-314-4532

## جدول نرمال ۱

■ مثال. جدول ALL\_SALES که اطلاعات فروش را نگهداری می کند در نظر بگیرید. آیا در فرم نرمال اول هست ؟

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
12351	Aug 14 2002	AQX88916	55	1317.25	Dave Williams	6794-1674	Diane	Adams	364 East Road	150

# جدول نرمال ۱

این جدول در فرم اول نرمال هست چون هیچ کدام از ستون ها چندمقداری نیستند بنابراین نیازی نیست روی جدول کاری انجام دهیم بجز اینکه یک کلید انتخاب نمائیم.

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
12351	Aug 14 2002	AQX88916	55	1317.25	Dave Williams	6794-1674	Diane	Adams	364 East Road	150

# در پایگاه داده (Functional Dependency) وابستگی تابعی

■ **وابستگی تابعی (FD)** صفت  $Y$  با صفت  $X$  وابستگی تابعی دارد اگر و فقط اگر در طول حیات رابطه به هر مقدار  $X$  دقیقا یک مقدار از  $Y$  متناظر باشد که میگوییم صفت  $X$  صفت  $Y$  را تعیین می کند.

■ **تعریف:** اگر صفت  $X$  صفت  $Y$  را تعیین کند، گفته می شود  $Y$  به صورت **تابعی وابسته** به  $X$  است و به صورت زیر نمایش داده می شود:

$$X \rightarrow Y$$

■ در اینجا:

•  $X$  **تعیین کننده (Determinant)** نام دارد.

•  $Y$  **وابسته (Dependent)** نام دارد.

شماره	نام	فامیل
۱۱	اکبر	حسینی
۲۲	اکبر	کریمی



# در پایگاه داده (Functional Dependency) وابستگی تابعی

FD1: StudentID  $\rightarrow$  StudentName ■

با دانستن StudentID ، می توانیم به طور منحصر به فرد StudentName را تعیین کنیم.

FD2: StudentID  $\rightarrow$  Course ■

با دانستن StudentID ، می توانیم Course را تعیین کنیم.

StudentID	StudentName	Course
101	آلیس	ریاضی
102	باب	فیزیک
103	چارلی	شیمی

اگر B زیرمجموعه A باشد،  $A \rightarrow B$  را وابستگی تابعی بدیهی می نامیم .

مثال:

$(Sname, avg) \rightarrow avg$

# در پایگاه داده (Functional Dependency) وابستگی تابعی

- وابستگی تابعی کامل FFD

- صفت خاصه  $Y$  به صفت خاصه  $X$  وابستگی تابعی کامل دارد اگر  $Y$  به  $X$  وابسته باشد ولی با هیچ یک از زیرمجموعه‌های  $X$  وابستگی تابعی نداشته باشد.

- رابطه دانشجوی زیر را در نظر بگیرید

- $S\#, SNAME, City, AVG, Department$

- وابستگی تابعی زیر وجود دارد

- $(S\#, SNAME) \rightarrow City$

- آیا وابستگی تابعی کامل هست؟

- اگر برای تمامی صفت‌های  $B$  داشته باشیم  $A \rightarrow B$  انگاه  $A$  ابر کلید هست.

## در پایگاه داده (Functional Dependency) وابستگی تابعی

- اگر  $F$  یک مجموعه از وابستگی های تابعی باشد آنگاه مجموعه تمام وابستگی های تابعی که از آن منتج می شود را مجموعه پوششی  $f$  می نامیم و با  $f^+$  نمایش می دهیم .

## قواعد استنتاج آرمسترانگ

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

# کاربردهای قواعد آرمسترانگ

■ (۱) محاسبه بستار صفت  $A^+$  :

● مجموعه تمام صفاتی که با  $A$  ، وابستگی تابعی دارند

● نکته: اگر  $A^+ = H_R$  در این صورت  $A$  سوپر کلید (الگوریتم تشخیص سوپر کلید)

■ (۲) محاسبه بستار مجوعه وابستگی‌های تابعی یک رابطه :  $F^+$

■ مجموعه تمام FD هایی که از  $F$  منطقاً استنتاج میشوند

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

■

# Finding $F^+$

■ اگر رابطه ی  $R(A,B,C,D)$  با وابستگی های تابعی  $F$  را داشته باشیم.

1.  $A \rightarrow B$

2.  $B \rightarrow C$

3.  $A \rightarrow D$

■ پیدا کردن  $F^+$  - بستار مجموعه شامل تمام وابستگی های تابعی مشتق شده است.

1.  $A \rightarrow B$

2.  $B \rightarrow C$

3.  $A \rightarrow D$

4.  $A \rightarrow CA$  (تعدی)

5.  $A \rightarrow BCA$  (افزایش)

6.  $A \rightarrow BCDA$  (ترکیب وابستگی)

# (2NF) Second Normal Form

## ■ تعریف 2NF

1. جدول باید ابتدا در 1NF باشد

2. تمام صفات غیر کلیدی باید به طور کامل به کل کلید اصلی وابسته باشند (نه فقط بخشی از کلید اصلی). اگر وابستگی جزئی وجود داشته باشد، جدول به 2NF نمی‌رسد.

ستون  $Y$  با ستون  $X$  در یک رابطه وابستگی تابعی (functional dependency) دارد اگر و فقط اگر به ازای هر مقدار در  $X$  دقیقا یک مقدار در  $Y$  متناظر با آن وجود داشته باشد. که به صورت  $X \rightarrow Y$  نشان داده می‌شود.

## جدول نرمال ۲

آیا جدول زیر 2NF هست؟

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
12351	Aug 14 2002	AQX88916	55	1317.25	Dave Williams	6794-1674	Diane	Adams	364 East Road	150



## جدول نرمال ۲

با کلید اصلی ترکیبی (SaleNo, ProductNo, CustomerNo)، وابستگی صفات غیر کلیدی را بررسی می‌کنیم:

1. Credit Limit: Address.Last.First

این صفات فقط به CustomerNo وابسته هستند و به کل کلید ترکیبی (SaleNo, ProductNo, CustomerNo) وابسته نیستند.

این موضوع باعث ایجاد وابستگی جزئی می‌شود و فرم ۲ NF را نقض می‌کند.

2. Salesrep: فقط به SaleNo وابسته است (با فرض اینکه هر فروش نماینده فروش خاصی دارد) و به کل کلید ترکیبی (SaleNo, ProductNo, CustomerNo) وابسته نیست.

این نیز یک وابستگی جزئی است که فرم ۲ NF را نقض می‌کند.

3. Amount به ProductNo و به Qty وابسته است، بنابراین به درستی به بخشی از کلید وابسته است. این مورد فرم ۲ NF را نقض نمی‌کند.

توجه کنید اگر کلیدهای کاندید در جدول، ترکیبی نباشند یعنی تنها شامل یک ستون باشند بلافاصله می‌گوئیم جدول NF۲ است.

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
					Dave Williams					

مثال. جدول ALL\_SALES را در نظر بگیرید:

ALL\_SALES(SaleNo, ProductNo, CustomerNo, SaleDate, QtyInStock, Description, Price, Customer\_Name, CreditLimit, Amount, Salesrep)

مشاهده می شود بعضی از ستون ها بهم مرتبط هستند و توسط بخشی از کلید مشخص می شوند.  
به عبارت دیگر بعضی ستون ها با زیرمجموعه ای از کلید وابستگی تابعی دارند:

ProductNo → {Description, ReorderLevel, Price, QtyInStock}

CustomerNo → {Customer\_Name, CreditLimit}

SaleNo → {Date, CustomerNo, ProductNo, Qty, Amount, Salesrep}

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
12351	Aug 14 2002	AQX88916	55	1317.25	Dave Williams	6794-1674	Diane	Adams	364 East Road	150

با جدا کردن این ستون ها به جداول جداگانه به فرم دوم نرمال می رسیم.

- PRODUCT(ProductNo, Description, Price, QtyInStock)
- CUSTOMER(CustomerNo, Customer\_Name, CreditLimit)
- SALE(SaleNo, Date, CustomerNo, ProductNo, Qty, Amount, Salesrep)

Sale No	SaleDate	ProductNo	Qty	Amount	Salesrep	Customer No	First	Last	Address	CreditLimit
12345	Aug 12 2002	AQX88916	1	23.95	Dave Williams	4649-4673	Richard	Johnston	14 West Avenue	1000
12346	Aug 12 2002	AQX88916	7	167.65	Sara Thompson	1113-7741	Wayne	Jones	42 York Street	<null>
12347	Aug 13 2002	AHL46785	3705	5001.75	Li Qing	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12348	Aug 13 2002	DHU69863	50	118.5	Sara Thompson	<null>	<null>	<null>	<null>	<null>
12349	Aug 14 2002	DHU69863	940	2227.8	Sara Thompson	1166-3461	Amelia	Waverley	995 Forth Street	<null>
12350	Aug 14 2002	DHU69863	42	99.54	Sara Thompson	7671-3496	Antonio	Gonzales	55B Granary Lane	<null>
12351	Aug 14 2002	AQX88916	55	1317.25	Dave Williams	6794-1674	Diane	Adams	364 East Road	150

# جدول نرمال ۲

■ یک جدول نرمال ۲ است اگر:

- نرمال ۱ باشد
- در آن هیچ وابستگی جزئی به کلید اصلی وجود نداشته باشد. به عبارت دیگر، هیچ ویژگی جدول تنها به قسمتی از کلید اصلی وابستگی نداشته باشد.

S#	Name	Crs#	Cname	Unit	Grade	Term
7801	علی	1400	پایگاه داده	3	20	79-2
7801	علی	1500	ریاضی 1	3	10	80-1
7801	علی	1600	تجزیه و تحلیل	3	20	80-1
7902	عسل	1400	پایگاه داده	3	7	80-1
7902	عسل	1700	تربیت بدنی	1	20	80-1

بخشی از جدول فوق وابسته به Crs# است و بخش دیگر وابسته به S# است

## جدول نرمال ۳

- یک جدول در فرم سوم نرمال 3NF است اگر اولاً 2NF باشد، ثانياً کلیه صفات خاصه غیر کلید در جدول با کلید اصلی وابستگی تابعی غیر تعدی داشته باشند.
- وابستگی تعدی transitive dependency یک وابستگی تابعی غیر مستقیم است که در آن  $X \rightarrow Z$  است اگر  $X \rightarrow Y$  و  $Y \rightarrow Z$  باشد.
- در فرم سوم نرمال کلیه ستون های جدول مستقیماً توسط کلید اصلی مشخص می شوند.
- با حذف فیلدهائی که وابستگی مستقیم با کلید ندارند به فرم سوم نرمال می رسیم. برای این کار گروهی از ستون های جدول را که مقدارشان برای بیش از یک رکورد تکرار می شود را در جدول جداگانه ای قرار دهید.

# جدول نرمال ۳

یک جدول نرمال ۳ است اگر:

● نرمال ۲ باشد

● در آن هیچ وابستگی تعدی (وابستگی با واسطه) در آن وجود نداشته باشد. به عبارت دیگر، در آن هیچ ویژگی غیر کلیدی به ویژگی غیر کلیدی دیگر وابستگی تابعی نداشته باشد.

Prof#	Pname	LastDegree	LastDegreeName
7801	علی	2	کاردانی
7802	آرش	2	کاردانی
7803	عسل	4	فوق لیسانس

● برای آنکه این جدول نرمال ۳ شود به صورت زیر تبدیل میشود:

- Prof(Prof#, Pname, LastDegree)
- Degree(LastDegree#, LastDegreeName)

# جدول نرمال ۳

■ مثال. فرض کنید جدول PRODUCT به صورت زیر جزئیات تولید کننده هر محصول را دارا باشد:

- **PRODUCT(ProductNo, Description, ReorderLevel, Price, QtyInStock, SupplierCode, SupplierName, SupplierAddress)**

این جدول کلید اصلی تک ستونی دارد بنابراین ۲ NF است. اگر تولید کننده چندین محصول را تولید کند فیلدهای SupplierName و SupplierAddress برای هر محصول تکرار می شود زیرا وابستگی تعدی با کلید اصلی دارند.

**ProductNo → SupplierCode → {SupplierName, SupplierAddress}**

با حذف این ستون ها و تقسیم جدول به صورت زیر به نرمال سوم می رسیم. توجه کنید که SupplierCode در جدول PRODUCT به عنوان کلید خارجی باقی می ماند.

**PRODUCT(ProductNo, Description, ReorderLevel, Price, QtyInStock, SupplierCode)**  
**SUPPLIER(SupplierCode, SupplierName, SupplierAddress)**

# جداول نرمال BCNF

- یک جدول نرمال BCNF است اگر و تنها اگر کلید تعیین کننده های (Determinant) آن، کلید کاندیدا باشند. یعنی هر رابطه  $A \rightarrow B$  در جدول وجود داشته باشد  $A$  کلید کاندیدا باشد.



# جداول نرمال BCNF

- **BCNF (Boyce-Codd Normal Form)** یکی از سطوح بالای نرمال سازی در پایگاه داده است که برای کاهش انحرافات و مشکلات ناشی از تکرار داده‌ها طراحی شده است. BCNF در حقیقت یک شکل اصلاح شده از **3NF (Third Normal Form)** است که شرایط سختگیرانه‌تری را برای روابط در پایگاه داده وضع می‌کند.

## شرایط: BCNF

- یک رابطه (Table) در BCNF قرار دارد اگر برای هر وابستگی تابعی  $X \rightarrow Y$  در آن رابطه، یکی از شرایط زیر برقرار باشد:
  1. **X** باید یک سوپرکلید باشد، یعنی مجموعه‌ای از ویژگی‌ها (attributes) که می‌تواند به طور منحصر به فرد هر سطر را شناسایی کند.
  2. در غیر این صورت، باید وابستگی‌های غیر ضروری (جزئی) وجود نداشته باشد.

# جداول نرمال BCNF

فرض کنید یک جدول به نام **Student\_Course** داریم که اطلاعات مربوط به دانشجویان و دوره‌هایی که در آن‌ها ثبت‌نام کرده‌اند را ذخیره می‌کند: در این جدول، وابستگی‌های تابعی به شکل زیر هستند:

- 1. Student\_ID, Course\_ID → Instructor:** برای هر ترکیب منحصر به فرد از دانشجو و دوره، استاد مشخص است.
- 2. Course\_ID → Instructor:** هر دوره یک استاد ثابت دارد.

مشکل این است که در این جدول، **Course\_ID → Instructor** نشان‌دهنده یک وابستگی است که باید در BCNF تجزیه شود زیرا **Course\_ID** یک سوپرکلید نیست (این تنها بخشی از کلید ترکیبی است).

Student_ID	Course_ID	Instructor	Semester
1	101	Dr. A	Fall 2024
2	102	Dr. B	Spring 2024
3	101	Dr. A	Fall 2024

# تبدیل به جدول نرمال BCNF

1. جدول **Course** که اطلاعات دوره‌ها و استادها را ذخیره می‌کند:
2. جدول **Student\_Course** که اطلاعات مربوط به دانشجویان و دوره‌ها را نگهداری می‌کند:

در این حالت، هر جدول به BCNF رسیده است، زیرا در هیچ کدام از جداول، وابستگی تابعی وجود ندارد که ویژگی‌های چپ آن‌ها سوپرکلید نباشد.

Course_ID	Instructor
101	Dr. A
102	Dr. B

Student_ID	Course_ID	Semester
1	101	Fall 2024
2	102	Spring 2024
3	101	Fall 2024

# جداول نرمال BCNF

■ آیا BCNF هست؟

در جدول زیر دو کلید کاندیدای  $S\#+Field$  و  $S\#+Tutor$  و نیز داریم  $Tutor \rightarrow Field$  بنابراین میتواند به دو جدول تقسیم شود

S#	Field	Tutor
۷۸۰۱	مهندس کامپیوتر	مهندس رضائی
۷۸۰۱	ریاضی محض	آرش ریاضیدان
۷۸۰۱	هنر	گلناز هنردوست
۷۹۰۲	مهندس کامپیوتر	مجید رضائی
۷۸۰۳	مهندس کامپیوتر	پروین صبا

# جداول نرمال BCNF

در جدول زیر دو کلید کاندیدای S#+Field و S#+Tutor و نیز داریم Tutor->Field بنابراین میتواند به دو جدول تقسیم شود

S#	Field	Tutor
۷۸۰۱	مهندس کامپیوتر	مهندس رضائی
۷۸۰۱	ریاضی محض	آرش ریاضیدان
۷۸۰۱	هنر	گلناز هنردوست
۷۹۰۲	مهندس کامپیوتر	مجید رضائی
۷۸۰۳	مهندس کامپیوتر	پروین صبا

Field	Tutor
مهندس کامپیوتر	مهندس رضائی
ریاضی محض	آرش ریاضیدان
هنر	گلناز هنردوست
مهندس کامپیوتر	مجید رضائی
مهندس کامپیوتر	پروین صبا

S#	Tutor
۷۸۰۱	مهندس رضائی
۷۸۰۱	آرش ریاضیدان
۷۸۰۱	گلناز هنردوست
۷۹۰۲	مجید رضائی
۷۸۰۳	پروین صبا

# جدول نرمال ۴

■ **4NF (Fourth Normal Form)** یک مرحله از نرمال سازی در طراحی پایگاه داده‌ها است که به منظور حذف وابستگی‌های چند مقداری (**Multivalued Dependencies**) طراحی شده است.

• این نوع وابستگی‌ها در زمانی که یک ویژگی در یک رابطه به بیش از یک مقدار وابسته باشد و این وابستگی‌ها مستقل از سایر ویژگی‌ها باشند، بروز می‌کند.

## ■ تعریف 4NF

■ یک رابطه در 4NF قرار دارد اگر:

• ابتدا در 3NF باشد.

• هیچ وابستگی چند مقداری در آن وجود نداشته باشد.

# جدول نرمال ۴

- مثال. اگر مشتریانی با چند آدرس داشته باشیم (که در محیط تجارت عادی است)، در جدول CUSTOMER نمی توانیم چند ستون آدرس را اضافه کنیم چون تعداد آدرس های ممکن را نمی دانیم.
- بنابراین ناگزیر به اضافه کردن رکورد جدید برای هر آدرس مشتری هستیم که باعث تکرار و افزونگی داده می شود. زیرا CustomerNo دیگر تنها یک آدرس را معین نمی کند بلکه مجموعه ای از آدرس های را نشان می دهد به عبارت دیگر وابستگی چندمقداری دارد.
- با حذف چنین وابستگی هائی و تقسیم جدول به صورت زیر به فرم چهارم نرمال می رسیم.
  - CUSTOMER(CustomerNo, First, Last, CreditLimit)
  - CUSTOMER\_ADDRESS(CustomerNo, Address)
- حالا هر مشتری می تواند هر تعداد آدرسی را داشته باشد.
- وابستگی چندمقداری (**multivalued dependency**) به این معنی است که حضور رکوردهای معینی در جدول وجود رکوردهای معین دیگری را برساند.

# جدول نرمال ۴

## وابستگی چند مقداری: (Multivalued Dependency)

وابستگی چند مقداری زمانی اتفاق می‌افتد که یک ویژگی (یا مجموعه‌ای از ویژگی‌ها) به بیش از یک مقدار وابسته باشد و این وابستگی‌ها به طور مستقل از سایر ویژگی‌ها باشند. به عبارت دیگر، در این نوع وابستگی، یک مجموعه از مقادیر برای یک ویژگی به مقادیر دیگری وابسته است بدون اینکه وابستگی بین آنها وجود داشته باشد.

### مثال:

فرض کنید یک جدول به نام **Student\_Course\_Hobby** داریم که اطلاعات مربوط به دانشجویان، دوره‌های ثبت‌نامی آن‌ها و سرگرمی‌هایشان را ذخیره می‌کند:

Student_ID	Course_ID	Hobby
1	101	Football
1	102	Music
1	103	Painting
2	101	Football
2	104	Chess

در این جدول، **Student\_ID** به **Course\_ID** و **Hobby** وابسته است، اما وابستگی‌ها به طور جداگانه و مستقل از یکدیگر وجود دارند:

- یک دانشجو می‌تواند در چندین دوره ثبت‌نام کند.
- همان‌طور که یک دانشجو ممکن است چندین سرگرمی داشته باشد.



# جدول نرمال ۴

## تجزیه به 4NF

برای رسیدن به 4NF، باید جدول را به دو جدول جداگانه تقسیم کنیم:

1. جدول **Student\_Course** که اطلاعات مربوط به دانشجویان و

دوره‌هایشان را ذخیره می‌کند:

2. جدول **Student\_Hobby** که اطلاعات مربوط به دانشجویان و

سرگرمی‌هایشان را ذخیره می‌کند:

در این حالت، هر جدول به صورت مستقل اطلاعات را ذخیره می‌کند و دیگر وابستگی چند مقداری وجود ندارد.

Student_ID	Course_ID
1	101
1	102
1	103
2	101
2	104

Student_ID	Hobby
1	Football
1	Music
1	Painting
2	Football
2	Chess

# جدول نرمال ۴

1. تمام کلید

2. اگر اضافه کنیم کارمند ۱۰۰، جاوا را به المانی میگوید، افزونگی رخ میدهد.

emp #	Skill	lang
۱۰۰	برنامه نویسی جاوا	انگلیسی
۱۰۰	تجزیه و تحلیل شی گرا	انگلیسی
۱۰۱	برنامه نویسی دلفی	انگلیسی
۱۰۱	تجزیه تحلیل شی گرا	انگلیسی
۱۰۱	طراحی وب سایت	انگلیسی
۱۰۱	برنامه نویسی دلفی	آلمانی
۱۰۱	تجزیه و تحلیل شی گرا	آلمانی
۱۰۱	طراحی وب سایت	آلمانی

emp #	Skill
۱۰۰	برنامه نویسی جاوا
۱۰۰	تجزیه و تحلیل شی گرا
۱۰۱	برنامه نویسی دلفی
۱۰۱	تجزیه تحلیل شی گرا
۱۰۱	طراحی وب سایت

emp#	lang
۱۰۰	انگلیسی
۱۰۱	انگلیسی
۱۰۱	آلمانی

# جدول نرمال ۵

یک جدول نرمال ۵ است اگر:

- نرمال ۴ باشد.

- نتوان آنرا به جداول کوچکتر تجزیه کرد بطوریکه حداقل یکی از جداول شامل هیچ یک از کلیدهای کاندیدای جدول اولیه نباشد.

# جدول نرمال ۵

آیا  $sp^{\infty}pj$  جدول اول را میدهد؟

S#	P#	J#
S1	P1	J2
S1	P2	J1
S2	P1	J1
S1	P1	J1

تجزیه

S#	P#
S1	P1
S1	P2
S2	P1

SP

P#	J#
P1	J2
P2	J1
P1	J1

PJ

S#	J#
S1	J2
S1	J1
S2	J1

SJ

S#	P#	J#
S1	P1	J2
S1	P2	J1
S2	P1	J1
<b>S2</b>	<b>P1</b>	<b>J2</b>
S1	P1	J1

$sp^{\infty}pj$

## معايب نرمال سازي

- نرمال سازي تکنیک مهمی برای طراحی پایگاه داده های کارآمد است اما در ضمنی که افزونگی داده را کاهش می دهد زیرا:
  - سبب کاهش سرعت اجرای سیستم می شود.
  - درجات بالای نرمال معمولا جداول بیشتر را می طلبند.
  - برای پاسخ به پرس و جوها گاهی باید کلیه جداول تقسیم شده دوباره با هم الحاق شوند
- در کاربردهائی که زمان پاسخ مهم است (نظیر وب) مطلوب نیست.

## معایب نرمال سازی

■ بالاترین سطح نرمال سازی باید با توجه به عملیات کاربردی در نظر گرفته شود:

- در پایگاه داده هایی که **بیشتر خواندنی** هستند و افزودگی داده در آنها مشکل **حادی نیست**، مانند داده های کاتالوگ یک سایت تجارت الکترونیکی، می توان سطح نرمال سازی را کاهش داد. به این عمل **denormalization** می گویند.
- در کاربردهائی که **درگیر داده های مهم** مانند داده های مالی هستند که **دائما در حال تغییرند** و باید سازگار باقی بمانند، احتمالا سعی می شود به سطوح بالاتر نرمال برسند حتی اگر سرعت پایگاه داده کم شود.

## معایب نرمال سازی

- گاهی با توجه به وضعیت ممکن است داده ها از چند پایگاه داده نرمال شده استخراج شوند و در یک انبار داده غیر نرمال قرار گیرد.
- این روش برای مخزن داده **Data warehouse** استاندارد خوبی است.

# غیرنرمال سازی

■ سناریو: برنامه تجارت الکترونیک

■ فرض کنید یک پایگاه داده برای یک پلتفرم تجارت الکترونیک طراحی می کنید با ساختار عادی شده زیر:

1. جدول محصولات:

ProductID	ProductName	CategoryID
101	Laptop	1
102	Smartphone	2

2. جدول دسته بندی ها:

CategoryID	CategoryName
1	Electronics
2	Mobile Phones

3. جدول سفارش ها:

OrderID	Customer ID	ProductID	OrderDate
1	201	101	2024-12-01

■ کوئری در ساختار عادی شده:

■ برای تولید گزارشی از سفارش ها به همراه نام محصولات و دسته بندی آنها، به چندین اتصال نیاز دارید:

```
SELECT o.OrderID, o.CustomerID, p.ProductName,  
c.CategoryName, o.OrderDate  
FROM Orders o  
JOIN Products p ON o.ProductID = p.ProductID  
JOIN Categories c ON p.CategoryID = c.CategoryID;
```



# غیرنرمال سازی

- برای بهینه‌سازی عملکرد، می‌توانید داده‌های دسته‌بندی را مستقیماً در جدول محصولات ذخیره کنید و جزئیات محصول و دسته‌بندی را به جدول سفارش‌ها اضافه کنید:

ProductID	ProductName	CategoryName
101	Laptop	Electronics
102	Smartphone	Mobile Phones

OrderID	CustomerID	ProductName	CategoryName	OrderDate
1	201	Laptop	Electronics	2024-12-01

1. جدول محصولات غیرنرمال شده:

2. جدول سفارش‌ها غیرنرمال شده:

- مزایای غیرعادی سازی:

1. کوئری‌های ساده‌تر:

- `SELECT OrderID, CustomerID, ProductName, CategoryName, OrderDate FROM Orders;`

- این کوئری هیچ پیوندی ندارد.

2. بهبود عملکرد کوئری: از آنجا که اتصالات حذف شده‌اند، بازیابی داده‌ها سریع‌تر انجام می‌شود، به‌ویژه برای مجموعه داده‌های بزرگ در محیط‌های پرتراфик.

- معایب:

1. افزونگی: اگر نام دسته‌بندی تغییر کند مثلاً "Electronics" به "Electronic Devices"، باید در مکان‌های مختلف به‌روزرسانی شود.

2. احتمال ناسازگاری داده‌ها: اگر به‌درستی مدیریت نشود، غیرعادی‌سازی ممکن است منجر به داده‌های ناسازگار شود.

# پایان فصل نرمال سازی