**نرمال سازی پایگاه داده چیست؟**

**نرمال سازی پایگاه داده** یعنی فرآيند سازماندهی داده در پايگاه داده بطور کارآمد است. نرمال سازی روشی برای طراحی جداول [پايگاه داده](https://programstore.ir/%d9%be%d8%a7%d9%8a%da%af%d8%a7%d9%87-%d8%af%d8%a7%d8%af%d9%87-data-base/) است و داده ها را به روشی ساماندهی می کند که باعث **کاهش افزونگی داده** و رفع مشکلات ساختاری و آنومالی شود. هدف از نرمال سازی حذف افزونگی داده و باقی نگهداشتن وابستگی بين داده های مرتبط است. به اين طريق اندازه پايگاه داده را کاهش داده و ذخيره منطقی داده را تضمين می کند.

مفهوم نرمال سازی پايگاه داده اولين بار توسط [Edgar Frank Codd](https://en.wikipedia.org/wiki/Edgar_F._Codd) معرفی شد. فرآيند نرمال سازی شامل ايجاد جداول و برقراری ارتباط بين آنها طبق قواعد معين است و روی وابستگی های ستون های جدول تمرکز دارد. اين فرآيند اغلب باعث ايجاد جداول بيشتر می شود ولی با وجودي که اثر تکرار داده درون پايگاه داده را دارد باعث افزونگی غير ضروری داده نمی شود**. هدف از نرمالسازی تضمين اين است که ستون های غير کليدی در هر جدول مستقيماً وابسته به کل کليد باشند و به اين ترتيب حذف وابستگی های ناسازگار، کاهش افزونگی، آنومالی کمتر و بهبود کارائی را نتيجه می دهد.**

**فرم های نرمال در پایگاه داده**

تئوری پايگاه داده درجه نرمال سازی جدول را با اصطلاح فرم های نرمال (normal form) شرح می دهد. فرم های نرمال (يا بطور خلاصه NF) معياری برای تعيين درجه نرمال جدول دراختيار می گذارد. فرم های نرمال جداگانه روی هر جدول می توانند بکار بروند. پايگاه 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)

اگر فرم اول نرمال در جدولی مشاهده شود اصطلاحا آنرا در فرم اول نرمال (1NF) می نامند. اگر سه فرم اول نرمال ديده شود آن را در فرم سوم نرمال (3NF) درنظر می گيرند. جدولی که دارای فرم نرمال درجه بالاتر باشد فرم های نرمال درجه پائين تر را هم دارا می باشد. بنابراين مثلا اگر جدولی 3NF باشد 2NF و 1NF هم هست. ولی عکس اين صحت ندارد. توجه داشته باشيد که نرمال سازی يک فرآيند تکراری نيست. يک جدول ممکن است در يک مرحله به فرم سوم نرمال دربيايد. بعلاوه اگر 3NF باشد به احتمال بسيار زياد 5NF هم خواهد بود.

فرم نرمال هريک باعث کاهش بيشتر افزونگی و تقسيم جداول به واحدهای کوچکتر می شوند. سه فرم اول نرمال (1NF، 2NF و 3NF) در ابتدا توسط Codd تعريف شد که به طور خلاصه **وابستگی صفات خاصه غير کليد را به کليد الزام می کنند**. فرم های چهارم و پنجم (4NF و 5NF) با ارتباطات چند به چند و يک به چند بيت صفات خاصه سروکار دارند. دو فرم ديگر هم وجود دارد که کاملا با اين جريان جور نمی شوند که BCNF و DK/NF هستند.

**نرمال سازی First Normal Form) 1NF)**

يک جدول در فرم اول نرمال (1NF) است **اگر و فقط اگر فاقد گروه داده تکرار شونده باشد. به عبارت ديگر هر ستون در جدول دارای مقدار اتميک باشد.** در مدل رابطه ای هر جدولی حداقل در فرم اول نرمال هست زيرا از الزامات مدل اين است که هر جدول شامل دقيقا يک مقدار برای هر صفت خاصه باشد که اصطلاحا “فاقد گروه تکرار شونده” گفته می شود.

**نرمال سازی Second Normal Form) 2NF)**

يک جدول در فرم دوم نرمال (2NF) است **اگر اولا 1NF باشد و ثانيا کليه ستون های غيرکليد با کليد اصلی وابستگی تابعی کامل داشته باشند**. ستون Y با ستون X در يک رابطه وابستگی تایعی (functional dependency) دارد اگروفقط اگر به ازای هر مقدار در X دقيقا يک مقدار در Y متناظر با آن وجود داشته باشد. که به صورت X→Y نشان داده می شود.

**نرمال سازی Third Normal Form) 3NF)**

يک جدول در فرم سوم نرمال (3NF) است اگر اولا 2NF باشد، ثانيا کليه صفات خاصه غير کليد در جدول با کليد اصلی وابستگی تابعی غير تعدی داشته باشند. وابستگی تعدی (transitive dependency) يک وابستگی تابعی غير مستقيم است که در آن X→Z است اگر X→Y و Y→Z باشد.

در فرم سوم نرمال کليه ستون های جدول مستقيما توسط کليد اصلی مشخص می شوند. با حذف فيلدهائی که وابستگی مستقيم با کليد ندارند به فرم سوم نرمال می رسيم. برای اين کار گروهی از ستون های جدول را که مقدارشان برای بيش از يک رکورد تکرار می شود را در جدول جداگانه ای قرار دهيد.

**نرمال سازی Forth Normal Form) 4NF)**

يک جدول در فرم چهارم نرمال (4NF) است اگر اولا 3NF باشد، ثانيا هيچ ستونی در جدول وابستگی چند مقداری نداشته باشد. وابستگی چندمقداری (multivalued dependency) به اين معنی است که حضور رکوردهای معينی در جدول وجود رکوردهای معين ديگری را برساند.

**مثال عملی نرمال سازی پایگاه داده**

فرض کنید یک فروشگاه رسانه های تصویری یک بانک اطلاعاتی از فیلم های اجاره ای را ذخیره می کند. بدون هرگونه نرمال سازی ، تمام اطلاعات مطابق شکل زیر در یک جدول ذخیره می شوند.



در اینجا می بینید ستون نام فیلم دارای مقادیر مختلفی است.

**فرم های نرمال پایگاه داده**

حال به فرم نرمال سازی 1NF می رویم

**قوانین فرم نرمال سازی First Normal Form) 1NF)**

* هر سلول جدول باید دارای یک مقدار واحد باشد.
* هر رکورد باید منحصر به فرد یا unique باشد.

جدول فوق در 1NF

**مثال 1NF**



قبل از هر چیز چند نکته را باید درک کنیم

**کلید چیست؟**

کلید یا KEY یک مقدار است که برای شناسایی یک رکورد در یک جدول به صورت منحصر به فرد استفاده می شود. یک کلید می تواند یک ستون یا ترکیبی از ستون های متعدد باشد

توجه: ستون های جدول که برای شناسایی یک رکورد به طور منحصر به فرد استفاده نمی شوند ، ستون های غیر کلیدی نامیده می شوند.

**کلید اصلی یا Primary Key چیست؟**

کلید اصلی یک مقدار ستون واحد است که برای شناسایی یک رکورد پایگاه داده به صورت منحصر به فرد استفاده می شود که خصوصیات زیر را دارد:

* یک کلید اصلی نمی تواند NULL یا تهی باشد
* مقدار اصلی کلید باید منحصر به فرد باشد
* مقادیر اصلی کلید به ندرت تغییر می یابد
* با وارد کردن یک رکورد جدید ، به کلید اصلی باید مقداری داده شود

**کلید مرکب یا Composite چیست؟**

یک کلید مرکب یا کامپوزیت یک کلید اصلی است که از ستون های مختلفی تشکیل شده است که برای شناسایی یک رکورد به صورت منحصر به فرد استفاده می شوند

در بانک اطلاعاتی ما ، دو نفر با نام **سحر عبادی** داریم ، اما آنها در جاهای مختلف زندگی می کنند.



از این رو ، ما به نام کامل و آدرس نیاز داریم تا یک رکورد را بطور منحصر به فرد شناسایی کنیم. این یک کلید مرکب است.

**قوانین نرمال سازی Second Normal Form) 2NF)**

* قانون 1 : در 1NF باشد
* قانون 2 : کلید اصلی تک ستونی

واضح است که ما نمی توانیم به جلو حرکت کنیم تا بتوانیم بانک اطلاعاتی ساده خود را در فرم نرمال سازی 2NF ایجاد کنیم مگر اینکه جدول فوق را تقسیم کنیم.



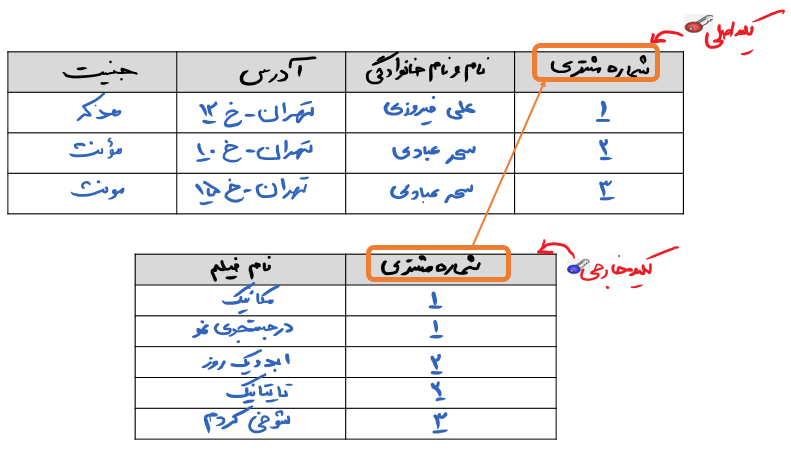
جدول 1NF ما به دو جدول تقسیم شده است. جدول 1 و جدول 2. جدول 1 حاوی اطلاعات عضو است. جدول 2 حاوی اطلاعاتی درباره فیلم های اجاره ای است.

ستون جدیدی به نام **شماره مشتری** را معرفی کرده ایم که کلید اصلی جدول 1 است. سوابق با استفاده از شناسه عضویت می توانند منحصراً در جدول 1 مشخص شوند.

**کلید خارجی یا Foreign Key چیست؟**

در جدول 2 بالا شماره مشتری یک کلید خارجی است. کلید اصلی خارجی کلید اصلی جدول دیگری را ارجاع می دهد یعنی به اتصال جداول شما کمک می کند

* یک کلید خارجی می تواند از کلید اصلی خود نام دیگری داشته باشد
* این اطمینان حاصل می کند که ردیف ها در یک جدول دارای سطرهای مربوطه در دیگری هستند
* بر خلاف کلید Primary ، آنها منحصر به فرد نیستند. اما اغلب آنها اینگونه نیستند
* کلیدهای خارجی می توانند تهی باشند.



**قوانین فرم نرمال سازی (3NF (Third Normal Form**

* قانون 1 : در 2NF باشد
* قانون 2 : هیچ وابستگی عملکردی انتقالی ندارد

برای انتقال جدول 2NF ما به 3NF ، باید دوباره جدول خود را تقسیم کنیم.



ما دوباره جداول خود را تقسیم کرده ایم و یک جدول جدید ایجاد کرده ایم که در آن نوع جنسیت ها ذخیره می شود. هیچ وابستگی عملکردی انتقالی وجود ندارد ، و از این رو جدول ما در 3NF است. در جدول 3 شناسه سلامتی کلید اصلی است ، و در جدول 1 شناسه سلامتی کلید اصلی اصلی در جدول 3 است.

اکنون نمونه کوچک ما در سطحی است که بیشتر برای دستیابی به اشکال بالاتر عادی سازی قابل تجزیه نیست. در واقع ، اینک در اشکال نرمال سازی بالاتر قرار دارد. تلاش های جداگانه برای انتقال به سطوح بعدی نرمال سازی داده ها معمولاً در پایگاه های داده پیچیده مورد نیاز است.

**معایب نرمال سازی**

نرمال سازی پایگاه داده تکنيک مهمی برای طراحی پايگاه داده های کارآمد است اما در ضمنی که افزونگی داده را کاهش می دهد سبب کاهش اجرای سيستم می شود. درجات بالای نرمال معمولا جدوال بيشتر را می طلبند. برای پاسخ به پرس و جوها گاهی بايد کليه جداول تقسيم شده دوباره با هم الحاق شوند که در کاربردهائی که زمان پاسخ مهم است (نظير وب) مطلوب نيست.

بالاترين سطح نرمال سازی با توجه به عمليات کاربردی درنظر گرفته می شود. در پايگاه داده هايی که بيشتر خواندنی هستند و افزونگی داده در آنها مشکل حادی نيست، مانند داده های کاتالوگ يک سايت تجارت الکترونيکی، می توان سطح نرمال سازی را کاهش داد. به اين عمل denormalization می گويند. از طرف ديگر در کاربردهایی که درگير داده های مهم مانند داده های مالی هستند که دائما در حال تغييرند و بايد سازگار باقی بمانند، احتمالا سعی می شود به سطوح بالاتر نرمال برسند حتی اگر سرعت پايگاه داده کم شود.

گاهی با توجه به وضعيت ممکن است داده ها از چند پايگاه داده نرمال شده استخراج شوند و در يک انبار داده غير نرمال قرار گيرد. اين روش برای مخزن داده Data warehouse استاندارد خوبی است.