


بیت الله

ایجاد بانک های اطلاعاتی

مدرس: ابوالقاسم حسن پور

Database Systems



مدرس: ابوالقاسم حسن پور

مطالعه ی بسیار و پی گیر در مسائل علمی ، باعث شگفتنی
عقل و تقویت نیروی فکر و فهم است. امام صادق (ع)

مدرس: ابوالقاسم حسن پور

فصل چهارم

نرمالسازی رابطه ها

مدرس: ابوالقاسم حسن پور

نرمالسازی

- نرمالسازی فرآیند سازماندهی داده در پایگاه داده بطور کارآمد است. نرمالسازی روشی برای طراحی جداول پایگاه داده است و داده ها را به طریقی ساماندهی می کند که باعث کاهش افزونگی داده و رفع مشکلات ساختاری و آنومالی شود. هدف از نرمالسازی حذف افزونگی داده و باقی نگهداشتن وابستگی بین داده های مرتبط است. به این طریق اندازه پایگاه داده را کاهش داده و ذخیره منطقی داده را تضمین می کند.
- هدف از نرمالسازی تضمین این است که ستون های غیر کلیدی در هر جدول مستقیما وابسته به کل کلید باشند و به این ترتیب حذف وابستگی های ناسازگار، کاهش افزونگی، آنومالی کمتر و بهبود کارائی را نتیجه می دهد.

مدرس: ابوالقاسم حسن پور

نرمالسازی

- یک روش رایج و منطقی پایگاه داده ، تبدیل نمودار ER به مجموعه ای از رابطه هاست که اگر مدلسازی داده ها و فرآیند تبدیل نمودار به درستی انجام شود، معمولا مجموعه ای از رابطه های نرمال، خوش طرح داریم. با نرمالسازی طراحی منطقی پایگاه داده را بهبود می بخشیم.
- یک پایگاه داده را نرمال گویند اگر هر فیلد تنها یک مقدار بگیرد و هر جدول دارای مجموعه صحیح فیلدها باشد. صورت های نرمال عبارتند از:
 - 1NF- 2NF- 3NF- BCNF- 4NF - 5NF

مثال

از بانک اطلاعاتی عرضه کنندگان، قطعات و ارسال به عنوان مثال استفاده خواهد شد:

S { S# , Sname , Status , City }

Primary key { S# }

P { P# , Pname , Color , Weight , City }

Primary key { P# }

SP { S# , P# , QTY }

Primary key { S# , P# }

Foreign key {S#} References S

Foreign key {P#} References P

تجزیه بدون نقصان

یک متغیر رابطه ای شکسته یا تجزیه می شود و لازم است فرآیند تجزیه برگشت پذیر باشد، یعنی در این فرآیند هیچ اطلاعاتی نباید از بین برود.

متغیر رابطه ای عرضه کننده S را در نظر بگیرید. حال رابطه را تجزیه می کنیم.

S#	SName	Status	City

S#	Sname	City

S#	Sname	Status

۱- افزونگی داده‌ها (Data Redundancy)

۲- بی نظمی (Anomaly)

۳- مقادیر تهی (NULL Values)

سه اشکال عمده در
رابطه با جداول

مشکلات افزونگی

1. هدر رفتن فضای ذخیره سازی: با وجودیکه امروزه دیسک های چندصد گیگا بایتی وجود دارد چندین بار ذخیره یک داده غیر ضروری است.
2. آنومالی در بهنگام سازی: اگر داده یک مشتری، مثلاً آدرس، تغییر کند باید در همه جاهایی که ذخیره شده است این تغییر اعمال شود در غیر اینصورت جامعیت نقص می شود.
3. 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.
4. آنومالی در درج: به همین صورت نمی توانیم مشخصات مشتری جدید را درج کنیم مگر اینکه کالائی خریده باشد.

مدرس: ابوالقاسم حسن پور



افزونگی داده ها (Data Redundancy)

۱- افزونگی داده‌ها (Data Redundancy)

- افزونگی یعنی تکرار بی‌رویه داده‌ها.

- در بانک اطلاعات رابطه‌ای، تکرار داده‌ها تنها راه برقراری ارتباط بین جداول است و از آن به عنوان کلید خارجی یاد می‌شود. تکرار بیش از این، افزونگی نام دارد.

مدرس: ابوالقاسم حسن پور



مقادیر تهی (Null Values)

۳- مقادیر تهی (NULL Values)

- با ادغام جداول، از نشان دادن بعضی از اقلام اطلاعاتی، بدون استفاده از مقادیر تهی، ناتوانیم.
- مثال: معمولاً دانشکده‌ها همه درس‌های خود را در همه ترم‌ها ارائه نمی‌دهند. این درس‌ها یا باید در جدول نیابند یا برای سایر اطلاعات آن مقدار تهی در نظر بگیریم.

مدرس: ابوالقاسم حسن پور



بی‌نظمی (Anomaly)

۲- بی‌نظمی (Anomaly)

- وجود افزونگی باعث بی‌نظمی در تغییر داده‌ها می‌شود.

- بعنوان مثال اگر بخواهید یک نمره در جدول Term وارد نمایید، باید اطلاعات کامل درس را نیز همراه آن وارد نمایید. همین اتفاق در حذف کردن و به روز درآوردن داده‌ها نیز اتفاق می‌افتد. در این حالت کار کنترل دشوار می‌شود.

مدرس: ابوالقاسم حسن پور




وابستگی تابعی

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

وابستگی تابعی

$A \rightarrow B$

برقرار است اگر برای تمام رابطه‌ها در R ، به ازای هر مقدار A فقط یک مقدار در B داشته باشیم.

مدرس: ابوالقاسم حسن پور  13

مثال


• نکته: وابستگی تابعی باید از ذات صفات (معنای صفات) ناشی شده باشد.

• مثال: آیا $S\# \rightarrow avg$ برقرار است؟
• برقرار نیست.

• مثال: آیا $avg \rightarrow S\#$ برقرار است؟
• برقرار است.

• مثال: آیا $city, avg \rightarrow sname$ برقرار است؟
• برقرار نیست.

• مثال: $s\#, sname \rightarrow city, avg, clg\#$
• برقرار است.

مدرس: ابوالقاسم حسن پور  14

مثال

• نکته: وابستگی تابعی برای تعریف محدودیت‌های بانک اطلاعات نیز به کار می‌رود.

• مثال: $S\# \rightarrow avg$ ؟
• هر دانشجو یک معدل دارد. (یک محدودیت است)

• مثال: $sname \rightarrow s\#$ ؟
• درست نیست. دو نفر همتام نمی‌توانیم داشته باشیم.

مدرس: ابوالقاسم حسن پور  15


ابرقید

• اگر B شامل تمام صفات شمای R باشد و داشته باشیم $A \rightarrow B$ ، در اینصورت A را ابرکلید نامیده و می‌نویسیم $A \rightarrow R$.

• مثال: $s\# \rightarrow stud$ ؟

• مثال: $s\#, sname \rightarrow stud$ ؟

• مثال: $avg \rightarrow stud$ ؟

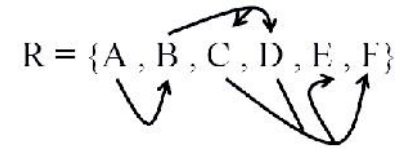
مدرس: ابوالقاسم حسن پور  16

• اگر $A \rightarrow R$ آنگاه گوییم، A کلید کاندیدای شمای R است.

• مثال: $s\#,sname \rightarrow stud$

$s\# \rightarrow stud$

کلید کاندید



• $F = \{A \rightarrow B, B \rightarrow CD, CD \rightarrow EF\}$

وابستگی تابعی

• آیا جداولی که طراحی کرده‌ایم بهترین است؟
• در مدل رابطه‌ای، با "نرمال‌سازی" می‌توان به این سؤال پاسخ داد.

• نرمال‌سازی یعنی پیروی از یک سری فرم‌های نرمال که منجر به تجزیه جداول می‌گردد.

نرمال‌سازی

- 1NF
- 2NF
- 3NF
- BCNF
- 4NF
- 5NF

فرم‌های نرمال

نکته: نرمال‌سازی تا سطح 3NF اجباری است.

1NF

- جدولی در 1NF است که:
 - صفات آن از دامنه ساده باشند. یعنی تنها دارای یک مقدار باشد.
 - کلیدهای اصلی و فرعی آن تعیین شده باشد.
 - تمام صفتهای آن به کلید اصلی وابستگی تابعی داشته شده باشند.
- مثال: ذخیره سازی دو شماره تلفن در یک فیلد مجاز نیست.



 مدرس: ابوالقاسم حسن پور

1NF

- $crs = \{c\#, cname, unit, clg\# \}$
- $clg = \{clg\#, clgname, city, pname \}$
- $sec = \{s\#, c\#, term, sec\#, pname, score \}$

این سه جدول در 1NF هستند.


مشکل 1NF وابستگی جزئی است یعنی وابستگی یک صفت غیر کلیدی به بخشی از کلید اصلی


 مدرس: ابوالقاسم حسن پور

1NF


سفرهای				اطلاعات مسیری			
شماره سفر	تعداد	شرح	قیمت	شماره	نام	تعداد	قیمت
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰

سفرهای				اطلاعات مسیری			
شماره سفر	تعداد	شرح	قیمت	شماره	نام	تعداد	قیمت
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰
۱۰۰	۳۰	راهن	۵۰	۱۰۰	خیابان خواجه نصیر	۱۰	۱۰۰


 مدرس: ابوالقاسم حسن پور

2NF

- 1NF باشد.
- وابستگی جزئی نداشته باشد. (صفاتهای آن به زیر مجموعه‌های کلید اصلی وابستگی نداشته باشند).


 مدرس: ابوالقاسم حسن پور

2NF

- جدول را به دو جدول به صورت زیر می‌شکنیم
۱. صفت یا صفاتی که ایجاد وابستگی جزئی کرده است با تمام وابسته‌هایش را در جدول اول قرار می‌دهیم .
 ۲. صفت یا صفاتی که ایجاد وابستگی جزئی کرده است را به همراه باقیمانده صفات در جدول دوم قرار می‌دهیم.
- نکته: ممکن است نیاز باشد مراحل فوق را دوباره تکرار کنید.

2NF

$sec = \{s\#, c\#, term, sec\#, pname, score\}$
 $enroll = \{c\#, term, sec\#, pname\}$
 $grade = \{s\#, c\#, term, sec\#, score\}$

مشکل فرم نرمال 2NF وابستگی انتقالی است یعنی وابستگی یک صفت غیر کلیدی به صفت غیر کلیدی دیگر

2NF

ساعات		اطلاعات دانشجوی		شماره	
تعداد	نرخ قیمت	نشانی	نام	شماره	شماره
۱۰۰	مقدار	۱۰۰	علی خجایان	۱۱	۱۰۰۱
۲۰۰	توزنگار	۲	علی خجایان	۱۱	۱۰۰۲
۵۰	پاکن	۳	علی خجایان	۱۱	۱۰۰۳
۲۰۰	خودکار	۵	رضا خجایان	۱۱	۱۰۰۴
۵۰	پاکن	۳	علی خجایان	۱۱	۱۰۰۲
۱۰۰	مقدار	۳	علی خجایان	۱۱	۱۰۰۳

ساعات		شماره	
تعداد	نرخ قیمت	شماره	شماره
۱۰۰	مقدار	۱۰۰	۱۰۰۱
۲۰۰	توزنگار	۲	۱۰۰۲
۵۰	پاکن	۳	۱۰۰۳
۲۰۰	خودکار	۵	۱۰۰۴
۵۰	پاکن	۲	۱۰۰۲
۱۰۰	مقدار	۳	۱۰۰۳

شماره	نام	نشانی
۱۱	علی خجایان	۱۰۱
۱۲	رضا خجایان	۱۰۲
۱۱	علی خجایان	۱۰۱


3NF

- 2NF باشد.
- وابستگی انتقالی نداشته 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 را به دو جدول به صورت زیر می‌شکنیم

۱. صفت یا صفاتی که ایجاد وابستگی انتقالی کرده است با تمام وابسته‌هایش را در جدول اول قرار می‌دهیم.
۲. صفت یا صفاتی که ایجاد وابستگی انتقالی کرده است را به همراه باقیمانده صفات در جدول دوم قرار می‌دهیم.

3NF


مدرس: ابوالقاسم حسن پور  29

• $crs = \{ c\#, cname, unit, clg\# \}$

• $crs-pres = \{ cname, unit, clg\# \}$

• $crs1 = \{ c\#, cname, unit \}$


3NF

مدرس: ابوالقاسم حسن پور  30

• 3NF با جداولی که هر سه شرط زیر را دارند ممکن است مشکل داشته باشد.

۱. جدول دارای حداقل دو کلید کاندید باشد.
۲. این کلیدهای کاندید ترکیبی باشند.
۳. این کلیدهای کاندید ترکیبی، صفتهای مشترکی داشته باشند.

3NF


مدرس: ابوالقاسم حسن پور  31

سر	تعداد	نوع	قیمت
۱۰۰	۱۰	۱	۲۰۰
۲۰۰	۲۰	۲	۱۰۰
۳۰۰	۳۰	۳	۱۰۰
۴۰۰	۴۰	۴	۱۰۰
۵۰۰	۵۰	۵	۱۰۰
۶۰۰	۶۰	۶	۱۰۰
۷۰۰	۷۰	۷	۱۰۰

کلاس	تعداد	قیمت
۱۰۱	۱۰	۲۰۰
۱۰۲	۲۰	۱۰۰
۱۰۳	۳۰	۱۰۰

نام	نشانی	تلفن
علی	خیابان ۱۰۱	۱۱
رضا	خیابان ۱۰۲	۱۲

3NF

مدرس: ابوالقاسم حسن پور  32

- جدولی در BCNF است که ستون‌های آن فقط به کلیدهای کاندیدش وابستگی تابعی داشته باشند.
- برخلاف فرم‌های نرمال دیگر، BCNF بدون استفاده از 3NF و فرم‌های نرمال قبلی تعریف می‌شود.
- در مواردی، نرمال‌سازی تا سطح BCNF لازم نیست و بهتر است از تبدیل جدول فرم 3NF به BCNF خودداری کرد.

BCNF

مدرس: ابوالقاسم حسن پور 33

- یک جدول در فرم چهارم نرمال است اگر اولاً 3NF باشد، ثانیاً هیچ ستونی در جدول وابستگی چند مقداری نداشته باشد. وابستگی چندمقداری (multivalued dependency) به این معنی است که حضور رکوردهای معینی در جدول وجود رکوردهای معین دیگری را برساند.


4NF

مدرس: ابوالقاسم حسن پور 34

- نرمال‌سازی تکنیک مهمی برای طراحی پایگاه داده‌های کارآمد است اما در ضمنی که افزودنی داده را کاهش می‌دهد سبب کاهش اجرای سیستم می‌شود. درجات بالای نرمال معمولاً جدول بیشتر را می‌طلبند. برای پاسخ به پرس و جوها گاهی باید کلیه جداول تقسیم شده دوباره با هم الحاق شوند که در کاربردهایی که زمان پاسخ مهم است (نظیر وب) مطلوب نیست.
- بالاترین سطح نرمال‌سازی با توجه به عملیات کاربردی در نظر گرفته می‌شود. در پایگاه داده‌هایی که بیشتر خواندنی هستند و افزودنی داده در آنها مشکل حادی نیست، مانند داده‌های کاتالوگ یک سایت تجارت الکترونیکی، می‌توان سطح نرمال‌سازی را کاهش داد. به این عمل denormalization می‌گویند. از طرف دیگر در کاربردهایی که درگیر داده‌های مهم مانند داده‌های مالی هستند که دائماً در حال تغییرند و باید سازگار باقی بمانند، احتمالاً سعی می‌شود به سطوح بالاتر نرمال برسند حتی اگر سرعت پایگاه داده کم شود.

معایب نرمال‌سازی

مدرس: ابوالقاسم حسن پور 35



4NF

مدرس: ابوالقاسم حسن پور 36