پایگاه داده

واحد کار دوم

پس از مطالعه این واحد کار از فراگیر انتظار می‌رود که:

۱ – انواع مدل‌های پایگاه داده را بشناسد.
۲ – مفهوم پایگاه‌های داده رابطه‌ای را بداند.
۳ – انواع رابطه‌ها را بشناسد.
۴ – اصول طراحی پایگاه داده را بداند.
۵ – با مفهوم نرمال‌سازی آشنا باشد.
۶ – انواع مدل‌های نرمال اول، دوم و سوم را بشناسد.

زمان(ساعت)	
نظری	عملی
۸	۵

مقدمه

با مفاهیم کلی بانک اطلاعاتی، موجودیت، رکورد، فیلد و ... در درس بانک اطلاعاتی سال گذشته آشنا شده‌اید.

در این درس ضمن یادآوری مباحث گذشته، با مفاهیم پیشرفته‌تر بانک اطلاعاتی نیز آشنا خواهید شد. در حقیقت سیستم بانک اطلاعاتی یا پایگاه داده‌ها، سیستم کامپیوتری نگهداری اطلاعات است. پایگاه داده را می‌توان به یک قفسه الکترونیکی تشبیه کرد که اطلاعات را به‌طور منظم درون خود نگهداری می‌کند.

کاربران سیستم بانک اطلاعاتی می‌توانند اعمال مختلفی روی آن انجام دهند که نمونه‌هایی از این اعمال عبارتند از:

- افزودن فایل‌های جدید و خالی به پایگاه داده
- افزودن داده درون فایل‌هایی که از قبل ایجاد شده‌اند.
- بازیابی داده‌ها از فایل‌های موجود
- تغییر در داده‌های فایل‌های موجود
- حذف داده‌ها از فایل‌های موجود
- حذف فایل‌های پایگاه داده

۱–۲ مدل‌های پایگاه داده

مدل‌های پایگاه داده‌ها یک طرح و الگوی کاری برای کاربران پایگاه داده‌ها را در سطح منطقی مشخص می‌کنند و معمولاً سیستم‌های مدیریت پایگاه داده براساس مدل‌هایی که استفاده می‌کنند تقسیم‌بندی می‌شوند.

نکته به برنامه کامپیوتری که برای مدیریت داده‌ها و پاسخگویی به کاربران از طریق پایگاه داده به‌کار برده می‌شود سیستم مدیریت پایگاه داده (DBMS) گفته می‌شود.

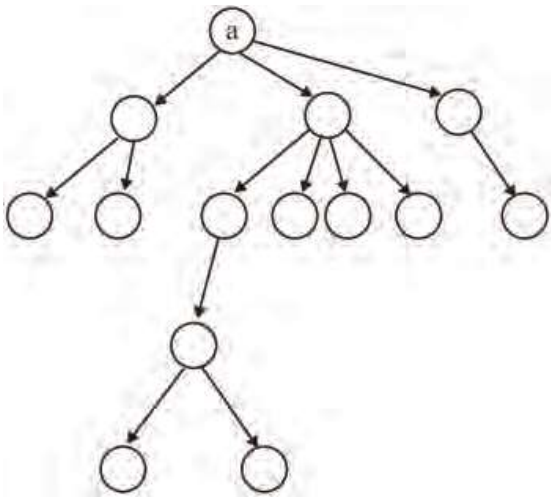
مهم‌ترین انواع مدل‌های پایگاه داده عبارتند از:

- پایگاه داده سلسله مراتبی
- پایگاه داده رابطه‌ای

- پایگاه داده شبکه‌ای
- پایگاه داده شئ‌گرا

۱-۱-۲ مدل سلسله مراتبی

این ساختار یکی از قدیمی‌ترین مدل‌های طراحی بانک اطلاعاتی است. در این مدل داده‌ها و ارتباط بین آن‌ها توسط اشاره‌گرهایی با یکدیگر مرتبط می‌شوند.

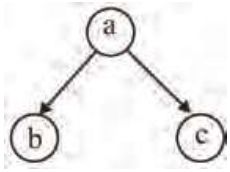


شکل ۱-۲

شکل ۱-۲ یک نمونه مدل سلسله مراتبی است. مدل سلسله مراتبی که مشابه ساختار درخت پیاده می‌شود، دارای یک گره ریشه a و Nگره فرزند است.

ارتباط بین گره‌ها از بالا به پایین صورت می‌گیرد و مسیری از گره‌های پایین‌تر به گره‌های بالاتر وجود ندارد.

هر گره پدری می‌تواند چند گره فرزند داشته باشد، اما هر گره فرزند، تنها یک گره پدر دارد.



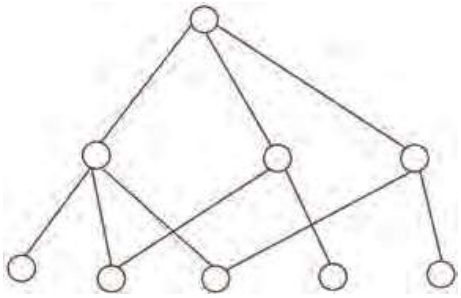
شکل ۲-۲

در شکل ۲-۲، a گره پدر و c و b گره‌های فرزند هستند.

در عمل پیاده‌سازی ساختار سلسله مراتبی بسیار مشکل است، به همین دلیل پس از مدتی جای خود را به مدل‌های دیگر پایگاه داده داد و عملاً منسوخ شده است.

۲-۱-۲ پایگاه داده شبکه‌ای

این مدل در سال ۱۹۶۹ و براساس دو مفهوم مهم مجموعه‌ها و رکوردها ارائه شد. پیاده‌سازی این مدل توسط گراف انجام می‌شود.



شکل ۳–۲

مزیت این روش این اســت که مدل‌های ارتباطی طبیعی بیشتری را بین موجودیت‌ها فراهــم می‌کنــدو در عین حــال پیچیدگی کمتری خواهد داشت.

اگر بخواهیم این مدل را توسط ساختار درختی پیاده کنیم، می‌توانیم بگوییم که هر گره فرزند می‌تواند بیش از یک گره پدر داشته باشد.

این مدل پیچیده‌تر از مدل سلســله مراتبی است و در نتیجه عملیات ذخیره و بازیابی پیچیده‌تری نیز دارد.

۳–۱–۲ مدل رابطه‌ای

پایگاه داده رابطه‌ای از دید کاربران شــامل یک مجموعه جدول اســت که هر جدول دارای ســطرها و ســتون‌هایی اســت. اطلاعات مربوط به هر ســتون، فیلدهای جدول و اطلاعات هر سطر مربوط به یکی از رکوردهای جدول است.

جدول ۱–۲

نام	نام‌خانوادگی	شماره دانش‌آموزی
علی	اقبالی	۱۳۴۵۷۹۳۲
مریم	احمدی	۳۲۴۳۳۵۹۲
میثم	بهرامی	۲۴۱۲۳۷۱۴
زهرا	فدایی	۱۲۳۹۲۳۴۳

در جدول ۱–۲، سه فیلد با عناوین نام، نام‌خانوادگی و شماره دانش‌آموزی وجود دارد و همچنین هر سطر اطلاعات مربوط به یکی از رکوردهای جدول را نمایش داده است.

فیلد: کوچک‌ترین واحد داده ذخیره شده در بانک اطلاعاتی است. جدول ۱–۲ شامل سه فیلد ایجاد شده با عناوین «نام»، «نام‌خانوادگی» و «شماره دانش‌آموزی» است.

رکورد: مجموعه‌ای از فیلدهای ذخیره شده مرتبط به هم است. به عنوان مثال در جدول ۱–۲ نام‌خانوادگی فدایی مربوط به فردی با نام زهرا و شماره دانش‌آموزی ۱۲۳۹۲۳۴۳ است و چنانچه این فرد از جدول حذف شود، تمام ویژگی‌های مربوط به او نیز از جدول حذف

می‌شـــوند و تغییر در مشـــخصات این فرد تأثیری در مشـــخصات افراد دیگر نخواهد داشت لذا هر سطر از جدول ۱-۲ به عنوان یک رکورد در نظر گرفته می‌شود.

مدل‌های رابطه‌ای چند ویژگی مهم دارند:

- ردیف‌های تکراری در آن‌ها وجود ندارد.
- ترتیب ردیف‌ها اهمیتی در ساختار مدل ندارد.
- ترتیب ستون‌ها اهمیتی در ساختار مدل ندارد.
- مقادیر هر فیلد غیرقابل تجزیه است.

مفهوم کلید

- کلید در مدل رابطه‌ای، صفت یا ستونی است که برای هر کدام از رکوردها (سطرها) مقدار منحصر به فردی دارد. به عنوان مثال در جدول ۱-۲، ســـتون شـــماره دانش‌آموزی می‌تواند به عنوان فیلد کلید در نظر گرفته شود زیرا یقین داریم که برای هر رکورد منحصر به فرد است و مقدار این فیلد برای دو رکورد متفاوت از جدول یکسان نخواهد بود.

محیط عملیاتی

سازمان، مؤسسه یا نهادی را در نظر بگیرید که قصد داریم برای آن یک بانک اطلاعاتی ایجاد کنیم. در این صورت به این سازمان، مؤسسه یا نهاد، محیط عملیاتی گفته می‌شود.

مثال‌های زیر نمونه‌هایی از محیط عملیاتی هستند:

۱ - یک مدرسه
۲ - یک دانشگاه
۳ - یک بیمارستان
۴ - یک اداره یا سازمان دولتی
۵ - یک بانک
۶ - یک کتابخانه

تمرین:
چند محیط عملیاتی نام ببرید.

در هـــر محیـــط عملیاتی با توجه به نیـــاز، تعدادی از اجزای آن بـــه عنوان موجودیت انتخاب می‌شـــوند، به عنوان مثال در محیط مدرســـه موجودیت‌هایی با عنوان دانش‌آموز، معلم و درس داریم. در محیط عملیاتی بیمارســـتان موجودیت‌های بیمار، پزشـــک، اتاق، بخش و ... وجود دارد.

تمرین:

چه موجودیت‌هایی برای محیط‌های عملیاتی بانک و کتابخانه می‌توان انتخاب کرد؟

محیط عملیاتی مدرســـه را در نظر بگیرید، هر کدام از موجودیت‌های محیط عملیاتی ویژگی‌ها و صفات متعددی دارند که فقط برخی از آن‌ها را می‌توان به عنوان فیلد در نظر گرفت، به عنوان مثال یک دانش‌آموز می‌تواند صفات و ویژگی‌های زیر را داشته باشد:

- نام
- نام‌خانوادگی
- نام پدر
- نام مادر
- رنگ چشم
- قد
- وزن
- تاریخ تولد
- و ...

اما همان‌طور که مشاهده می‌کنید تمام این صفات در محیط عملیاتی مدرسه کاربردی نیستند، به عنوان مثال، رنگ چشم، قد، وزن، نام مادر و ... در محیط عملیاتی مدرسه به‌کار نمی‌روند. علاوه بر این در یک باشگاه ورزشی، یک بیمارستان یا هر محیط عملیاتی دیگر، با توجه به نیازها، صفات و ویژگی‌های متفاوتی در نظر گرفته می‌شوند.

به‌طـــور کلی در محیط عملیاتی مدرســـه می‌توان ویژگی‌های زیـــر را برای هر کدام از موجودیت‌ها در نظر گرفت.

ویژگی‌های موجودیت دانش‌آموز

۱ – نام (name)
۲ – نام‌خانوادگی (family)
۳ – رشته تحصیلی (major)
۴ – تاریخ تولد (date)
۵ – محل تولد (area)
۶ – نمره (grade)

ویژگی‌های موجودیت معلم

۱ – نام (tname)
۲ – نام‌خانوادگی (tfamily)
۳ – آدرس (taddress)
۴ – تلفن (ttel)

ویژگی‌های موجودیت درس

۱ – کد درس (cnumber)
۲ – نام درس (cname)
۳ – تعداد ساعات در طول هفته (hcourse)

هر کدام از این موجودیت‌ها، یک جدول در محیط عملیاتی مدرسه و هر کدام از ویژگی‌ها و صفات موجودیت‌ها، یک فیلد در جدول‌های بانک اطلاعاتی خواهند بود.

همان‌طور که پیش از این اشاره شد در مدل‌های رابطه‌ای نباید ردیف‌های تکراری وجود داشته باشد، برای روشن شدن مطلب کلاس خود را در نظر بگیرید، معلم شما برای صدا کردن دانش‌آموزان از کدام صفت دانش‌آموزان استفاده می‌کند، ممکن است پاسخ دهید نام‌خانوادگی، استفاده از نام‌خانوادگی برای صدا کردن دانش‌آموزان به این دلیل است که احتمال تکراری بودن کمتری دارد، تا اینجا می‌توان گفت نام‌خانوادگی به عنوان فیلد کلید در نظر گرفته شده است، اما اگر حتی یک مورد تکراری وجود داشته باشد و دو دانش‌آموز دارای نام‌خانوادگی مشابه باشند، به دلیل احتمال بروز اشتباه دیگر نمی‌توان از نام‌خانوادگی به تنهایی استفاده کرد، در این صورت معمولاً نام به دنبال نام‌خانوادگی استفاده می‌شود.

یعنی دو فیلد نام و نام‌خانوادگی به عنوان کلید استفاده می‌شوند.

• اگر کلید انتخاب شده شامل یک فیلد باشد، آن را فیلد کلید ساده می‌گویند.
• اگر کلید انتخاب شده شامل دو یا چند فیلد باشد، آن را فیلد کلید مرکب می‌گویند.

اما اگر قدری دقیق‌تر شویم، می‌بینیم که باز هم احتمال تکراری بودن نام و نام‌خانوادگی و این‌که دو نفر دارای نام و نام‌خانوادگی یکسان باشند نیز وجود دارد. اگر بخواهیم محیط عملیاتی مدرسه را طوری طراحی کنیم که احتمال تکراری بودن فیلدها را به صفر برسانیم بهتر است به جای ترکیب فیلدها یک فیلد جدید با نام شماره دانش‌آموزی به ویژگی‌های موجودیت دانش‌آموز اضافه کنیم، در این صورت اطمینان حاصل خواهد شد که هیچ دو دانش‌آموزی اطلاعات کاملاً یکسان ندارند و در نتیجه ایجاد سطر تکراری در جدول نخواهیم داشت.

برای تهیه جدول‌ها نکات زیر را در نظر بگیرید:

۱ – هر موجودیت مستقل را به عنوان یک جدول در نظر بگیرید.

• معلم
• درس
• دانش‌آموز

۲ – صفت‌های موجودیت‌ها به عنوان فیلدهای جدول‌ها در نظر گرفته می‌شوند.

جدول مربوط به دانش‌آموز

جدول ۲-۲ فیلدهای جدول دانش‌آموز (Student)

شماره دانش‌آموزی* (id)	نام (name)	نام‌خانوادگی (family)	تاریخ تولد (date)	محل تولد (area)	نمره (grade)

جدول مربوط به معلم

جدول ۲-۳ فیلدهای جدول معلم (teacher)

کد معلم* tcode	نام tname	نام‌خانوادگی tfamily	آدرس taddress	تلفن ttel

جدول مربوط به درس

جدول ۴-۲ فیلدهای جدول درس (course)

کد درس* cnumber	نام درس cname	تعداد ساعات در طول هفته hcourse

نکته

فیلدهایی که کنار آن‌ها علامت * درج شده است، فیلدهای کلیدی هستند.

در هـر محیط عملیاتـی موجودیت‌ها با یکدیگر مرتبط هسـتند. در طراحی و تحلیل بانک‌های اطلاعاتی، وجود این ارتباط و بررسی ماهیت آن، یکی از مراحل اصلی کار است.

در محیط عملیاتی مدرسه، هر دانش‌آموز در سال تحصیلی در چند درس ثبت‌نام می‌کند.

به کلمات «هر» و «چند» در ابتدای موجودیت‌های دانش‌آموز و درس دقت کنید، وجود این کلمات به این معناست که یک دانش‌آموز چند درس را انتخاب کرده است.

ماهیت ارتباط دانش‌آموز- درس، «یک به چند» است.

۲-۲ انواع ماهیت ارتباط

با توجه به توضیحات بخش محیط عملیاتی، به تناظر بین تعداد موجودیت‌ها، ماهیت ارتباط گفته می‌شود و این تناظر می‌تواند انواع زیر را داشته باشد:

- ارتباط یک به یک (۱:۱)
- ارتباط یک به چند (۱:N)
- ارتباط چند به چند (N:N)

اگر در محیط عملیاتی مدرسـه موجودیت‌های مدرسـه و مدیر را نیز در نظر داشـته باشیم، در ارتباط یک به یک، یک نمونه از موجودیت اول با یک نمونه از موجودیت دوم ارتباط دارد، به عنوان مثال هر مدرسه یک مدیر دارد.

در ارتبـاط یک بـه چند، یک نمونه از موجودیت اول با چند نمونه از موجودیت دوم ارتباط دارد و در مقابل هر نمونه از موجودیت دوم با یک نمونه از موجودیت اول ارتباط دارد، به عنوان مثال یک مدرسه چند معلم دارد و هر معلم در یک مدرسه تدریس می‌کند.

در ارتباط چند به چند، یک نمونه از موجودیت اول با چند نمونه از موجودیت دوم ارتباط دارد و همچنین یک نمونه از موجودیت دوم با چند نمونه از موجودیت اول مرتبط است.

مثـال: در محیط عملیاتی مدرسـه با در نظـر گرفتن موجودیت‌هـای درس، معلم و دانش‌آموز، ماهیت‌های ارتباط و تناظر بین موجودیت‌ها به‌صورت زیر خواهد بود:

۱ - **درس- معلم:** هر معلم چند درس ارائه می‌دهد و هر درس توسط یک معلم ارائه می‌شود لذا نوع ارتباط بین درس و معلم یک به چند خواهد بود.

۲ - **درس- دانش‌آمـوز:** هر دانش‌آموز در هـر دوره تحصیلی چند درس را انتخاب می‌کنـد و یک درس توسـط چند دانش‌آموز اخذ می‌شـود. لذا نـوع ارتباط بین درس- دانش‌آموز چند به چند است.

تمرین:
نوع ارتباط معلم- دانش‌آموز را از لحاظ ماهیت ارتباط بررسی کنید.

۱-۲-۲ تأثیر ماهیت ارتباط بر بانک اطلاعات

ماهیت ارتباط بین موجودیت‌ها می‌تواند در طراحی جدول‌ها تأثیرگذار باشـد و حتی ممکن است لازم شود، در انتخاب و طراحی جدول‌ها یک بازبینی کلی انجام شود.

برای ایجاد ساختار نهایی جدول‌های بانک اطلاعاتی باید نکات زیر را در نظر گرفت:

۱ - در ارتباط‌های ۱:۱ جدول‌ها در هم ادغام می‌شوند.

۲ - در ارتباط‌های ۱:N فیلد اصلی طرف ۱ به عنوان کلید خارجی به جدول طرف N افزوده می‌شود.

۳ - در ارتباط‌های N:N یک جدول جدید سـاخته می‌شود، طوری که شامل کلیدهای هر دو طرف باشد.

۲-۳ محیط عملیاتی تولیدکنندگان قطعات صنعتی

به منظور ارائه مثال بیشـتر، یک محیط عملیاتی جدید را در نظر می‌گیریم. مثلاً محیط عملیاتی تولیدکنندگان قطعات صنعتی در شهرهای مختلف. این محیط عملیاتی دارای دو

موجودیت اصلی «قطعه» و «تولیدکننده» است.

برای هرکدام از موجودیت‌ها می‌توان صفات (فیلدهای) زیر را در نظر گرفت:

موجودیت قطعه (جدول S)

۱ – شماره قطعه (S#) *

۲ – نام قطعه (Sname)

۳ – تعداد (Status)

۴ – شهر (City)

موجودیت تولیدکننده (جدول P)

۱ – شماره تولیدکننده (P #)*

۲ – نام تولیدکننده (Pname)

۳ – رنگ (Color)

۴ – وزن (Weight)

۵ – شهر (City)

هر قطعه ممکن است توسط چند تولیدکننده تولید شود و همچنین هر تولیدکننده، چند قطعه تولید می‌کند، در نتیجه ماهیت ارتباط این دو موجودیت چند به چند است.

همان‌طور که پیش از این اشاره شد، در ارتباط‌های چند به چند، باید یک جدول جدید ساخته شود که فیلدهای آن حداقل باید شامل فیلدهای کلید اصلی دو جدول اولیه باشند. در نتیجه یک جدول جدید شامل دو موجودیت شماره قطعه و شماره تولیدکننده خواهیم داشت:

جدول (SP)

۱ – شماره قطعه (S#)

۲ – شماره تولیدکننده (P#)

در این محیط عملیاتی تعداد قطعه تولید شده نیز می‌تواند به جدول SP افزوده شود. (Qty)

با توجه به توضیحات فوق سه جدول مربوط به این محیط عملیاتی به این صورت ترسیم می‌شوند:

جدول تولیدکننده (P)

جدول ۵-۲

شماره تولیدکننده (P#)	نام تولیدکننده (Sname)	تعداد خط تولید (Status)	شهر (City)

جدول قطعه (S)

جدول ۶-۲

شماره قطعه(S#)	نام قطعه (Sname)	رنگ (Color)	وزن (Weight)	شهر (City)

جدول تولیدکننده- قطعه (SP)

جدول ۷-۲

شماره قطعه (S#)	شماره تولیدکننده (P #)	تعداد قطعه تولید شده (Qty)

جداول فوق را با داده‌های نمونه زیر در نظر می‌گیریم:

جدول ۸-۲ تولیدکننده

P

P#	Pname	Status	City
P1	احمدی	۲۰	تهران
P2	رضایی	۱۰	کرمانشاه
P3	سلیمی	۳۰	کرمانشاه
P4	مهدوی	۲۰	تهران
P5	میثمی	۳۰	شیراز

جدول ۹-۲ قطعه

S

S#	Sname	Color	Weight	City
S1	پیچ	قرمز	۱۲	تهران
S2	شیر	سبز	۱۷	کرمانشاه
S3	لوله	آبی	۱۷	بوشهر
S4	لوله	قرمز	۱۴	تهران
S5	پیچ‌گوشتی	آبی	۱۲	کرمانشاه
S6	فازمتر	قرمز	۱۹	تهران

جدول ۱۰-۲ تولید کننده - قطعه

S P

S#	P#	Qty
S1	P1	۳۰۰
S1	P2	۲۰۰
S1	P3	۴۰۰
S1	P4	۲۰۰
S1	P5	۱۰۰
S1	P6	۱۰۰
S2	P1	۳۰۰
S2	P2	۴۰۰
S3	P2	۲۰۰
S4	P2	۲۰۰
S4	P4	۳۰۰
S4	P5	۴۰۰

برای این‌که با مفهوم جدول SP بیشـــتر آشنا شوید، سطر اول را در نظر بگیرید، مفهوم سطر اول این است که:

قطعه S۱ به میزان ۳۰۰ عدد توسط تولیدکننده P1، تولید شده است، یا این‌که تولیدکننده P1از قطعه S1 به میزان ۳۰۰ عدد تولید کرده است.

تمرین:
مفهوم سطرهای دوم و آخر را بیان کنید.

۲-۴ نرمال‌سازی بانک‌های اطلاعاتی

در محیط عملیاتی تولیدکنندگان قطعات صنعتی ســه جــدول S، P و SP وجود دارد، کاملاً مشخص است که در جدول‌های فوق فیلدهای مربوط به هر کدام، کاملاً ضروری هستند و به درستی انتخاب شده‌اند.

فرض کنید، بــه جای جدول SP در بانک اطلاعاتــی، جدول SCP را بــا چهار فیلد S#، City، P# و Qty طراحی کرده باشــیم، در این صورت مقادیر نمونه‌ای جدول به این صورت خواهد بود:

جدول ۲-۱۱

S#	City	P#	Qty
S1	تهران	P1	۳۰۰
S1	تهران	P2	۲۰۰
S1	تهران	P3	۴۰۰
S1	تهران	P4	۲۰۰
S1	تهران	P5	۱۰۰
S1	تهران	P6	۱۰۰
S2	کرمانشاه	P1	۳۰۰
S2	کرمانشاه	P2	۴۰۰
S3	کرمانشاه	P2	۲۰۰
S4	تهران	P2	۲۰۰
S4	تهران	P4	۳۰۰
S4	تهران	P5	۴۰۰

با دقت در جدول ۲-۱۱ مشاهده می‌شود کــه داده‌هــای موجــود در فیلــد City کاملاً تکــراری هســتند و از طریــق جدول‌های P و S نیز قابل دسترســی هســتند، به این عمل تکراری بودن داده‌ها که موجب اشغال فضای بیشــتر در پایگاه داده‌ها و جدول‌ها می‌شود، عمل افزونگی داده‌ها گفته می‌شود.

برای طراحی یک پایگاه داده استاندارد

باید دقت کرد که عمل افزونگی داده‌ها اتفاق نیفتد.

در طراحی یک بانک اطلاعاتی پس از تحلیل و بررسی موجودیت‌ها و در نظر گرفتن صفات هر کدام از آن‌ها لازم است جداول به شکل نرمال دربیایند، نرمال‌سازی (Normalization)، فرایندی است در رابطه با بانک‌های اطلاعاتی که با دو هدف عمده زیر انجام می‌شود:

کاهش افزونگی اطلاعات: به این معنی که اطلاعات فقط در یک مکان (جدول) ذخیره شوند نه این‌که چند بار در جدول‌های مختلف تکرار شوند و در تمام بانک با استفاده از روابط (Relationship) تعریف شده قابل دسترسی باشند.

حفظ یکپارچگی اطلاعات: به این معنی که اعمال تغییرات روی اطلاعات (نظیر ایجاد، بهنگام‌سازی و حذف) در یک مکان انجام شود و به دنبال آن آثار تغییرات در تمام بانک مشاهده گردد. برای روشن شدن مفهوم یکپارچگی به مثال زیر توجه نمایید:

فرض کنید یک بانک اطلاعاتی دارای دو موجودیت کتاب و نویسنده باشد. هر یک از موجودیت‌های فوق دارای صفت‌های (Attribute) مختص به خود می‌باشند. به عنوان نمونه موجودیت «کتاب» دارای صفت «نام نویسنده» و موجودیت «نویسنده» دارای صفت‌های متعددی نظیر «نام نویسنده»، «آدرس نویسنده» و ... باشد. در صورتی که در موجودیت «کتاب» یک رکورد ایجاد نماییم بدون این‌که نام نویسنده آن را در موجودیت «نویسنده» ایجاد کرده باشیم، مسلماً یک ناهنجاری به‌وجود خواهد آمد.

با توجه به اهداف فوق می‌توان گفت که فرایند نرمال‌سازی از ناهنجاری‌های به وجود آمده به دلیل بروز تغییرات در بانک اطلاعاتی جلوگیری خواهد نمود. با اعمال فرایند نرمال‌سازی، یک بانک اطلاعاتی کارا و مطمئن خواهیم داشت. فرایند نرمال‌سازی، فرم‌های متفاوتی دارد که انواع متداول آن به شرح ذیل می‌باشند:

- فرم اول نرمال (NF۱)
- فرم دوم نرمال (NF۲)
- فرم سوم نرمال (NF۳)

۱–۴–۲ فرم اول نرمال ۱(NF)

موجودیت یا جدولی در فرم نرمال اول است که هیچ کدام از فیلدهای آن قابل تقسیم نباشد،

بـــه عنوان مثال در محیط عملیاتی مدرســـه موجودیت دانش‌آمـــوز را در نظر بگیرید، چنانچه جدول مربوط به آن را به‌صورت زیر طراحی کرده باشید:

جدول ۱۲-۲

نام و نام‌خانوادگی	نام پدر	تاریخ تولد	محل تولد	معدل

حال جدول فوق با مقادیر نمونه‌ای به‌صورت زیر خواهد بود:

جدول ۱۳-۲

نام و نام‌خانوادگی	نام پدر	تاریخ تولد	محل تولد	معدل
مینا رضایی	احمد	۷۵/۳/۱	تهران	۱۹/۳۰
مریم توکلی	محمدرضا	۷۴/۷/۲	تهران	۱۹/۰۲
سارا بهرامی	مهدی	۷۵/۶/۳	مشهد	۱۸/۲۵
زهرا ساداتی	محمدسعید	۷۴/۱۲/۵	کرمانشاه	۱۸/۱۷
فاطمه خدابخشی	علی‌اکبر	۷۵/۶/۶	تهران	۱۹/۳۵

مقادیر فیلد اول جدول قابل تجزیه هستند به این معنا که اگر هر کدام از اسامی نام و نام‌خانوادگی را تقســیم کنیم، آنگاه دو بخش اطلاعاتی معنادار خواهیم داشـــت، به عنوان نمونه در رکورد اول، فیلد اول دارای مقدار «مینا رضایی" اســـت، می‌توان در این فیلد دو مقدار "مینا" و "رضایی" را استخراج نمود طوری که هر کدام از این اطلاعات معنادار باشند (جزء اول نام و جزء دوم، نام‌خانوادگی رکورد اول است). در این حالت می‌توان گفت که جدول در فرم نرمال اول نیست، زیرا مقادیر فیلدها قابل تجزیه هستند.

ســـؤال: فیلد تاریخ تولد نیز به اجزای روز، ماه و ســـال قابل تجزیه اســـت، آیا در این حالت نیز می‌توان گفت جدول در حالت نرمال نوع اول قرار ندارد؟

پاسخ: اصطلاح معنادار وابسته به نوع محیط عملیاتی است، در محیط عملیاتی مدرسه، فیلد تاریخ به عنوان یک فیلد مجزا و غیرقابل تجزیه در نظر گرفته می‌شود. همین مسأله را می‌توان در فیلد شـــماره تلفن نیز بررســـی کرد، هر شماره تلفن از ترکیب چند عدد در کنار هم تشکیل شده است، اگر اعداد را تجزیه کنیم، معنایی در محیط عملیاتی نخواهند داشت. در نتیجه در جدول ۱۳-۲ فقط فیلد اول قابل تجزیه است.

۲-۴-۲ فرم دوم نرمال (2NF)

برای روشـــن شـــدن مفهوم فرم نرمال باید مفهوم وابســـتگی تابعی را بشناسید، مفهوم وابستگی تابعی در بانک‌های اطلاعاتی مشابه مفهوم تابع در ریاضی است.

مفهوم تابع در ریاضی: در یک رابطه با نام R، متغیر X با Y وابستگی تابعی دارد، اگر به ازای هر مقدار X فقط و فقط یک مقدار Y وجود داشته باشد، در این صورت وابستگی تابعی X و Y به‌صورت $X \longrightarrow Y$ **شان داده می‌شود.**

رابطه R را مطابق جدول ۱۴-۲ برای سه متغیر X و Y و Z در نظر بگیرید:

جدول ۱۴-۲

X	Y	Z
x1	y1	z1
x2	y2	z3
x1	y1	z2
x3	y2	z2
x4	y3	z3

در رابطه R، می‌خواهیم وابستگی تابعی $X \longrightarrow Y$ و $Z \longrightarrow Y$را بررسی کنیم:

وابستگی تابعی $X \longrightarrow Y$ **برقرار است، زیرا به ازای مقادیر یکسان از** x1، **مقادیر یکســـانی در** y1 وجود دارد، به عنوان مثال ســـتون X را در نظر بگیرید، این ستون دارای دو مقدار تکراری x1 است، که هر دو مقدار آن با y1 از ستون Y متناظر است، در نتیجه به دلیل مساوی بودن طرف دوم رابطه، وابستگی تابعی $X \longrightarrow Y$ برقرار است.

وابســـتگی تابعی $Y \longrightarrow Z$ برقرار نیســـت، ســـتون Y دارای دو مقدار تکراری y1 و y2 است، مقدار y1 متناظر با z1 و z2 در ستون Z است، لذا به دلیل عدم تساوی این دو مقدار می‌توان گفت وابستگی تابعی $Z \longrightarrow Y$برقرار نیست.

نکته برای بررســـی برقراری وابســـتگی تابعی، حتی اگر یک نمونه یافت شود که وابستگی تابعی را نقض می‌کند، باید گفت که وابستگی تابعی برقرار نیست.

با توجه به مفهوم وابستگی تابعی می‌توان مفهوم فرم دوم نرمال را بررسی کرد:
موجودیتی در فرم دوم نرمال است که اولاً در فرم نرمال اول باشد، ثانیاً تمام فیلدهای آن وابستگی تابعی را با کلید اصلی نقض نکنند.

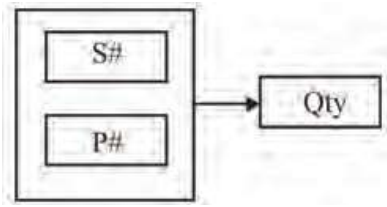
نکته اگر جدول دارای یک فیلد کلید ساده باشد، به دلیل ویژگی‌های فیلد کلید می‌توان گفت که جدول در فرم دوم نرمال قرار دارد، اما اگر یک جدول دارای فیلد کلید اصلی مرکب باشد، به این سادگی نمی‌توان نتیجه گرفت که جدول در حالت فرم دوم نرمال قرار دارد یا خیر.

جدول‌های مربوط به محیط عملیاتی تولیدکنندگان قطعات را در نظر بگیرید (جدول‌های ۸-۲، ۹-۲ و ۱۰-۲).

در جدول‌های S و P به دلیل نرمال نوع اول بودن و وجود کلید اصلی ساده می‌توان به راحتی نتیجه گرفت که رابطه مربوط به آن‌ها در حالت نرمال دوم قرار دارد.

اما در جدول SP، کلید اصلی شامل ترکیبی از فیلدهای S#و P# است، در نتیجه برای بررسی این‌که آیا در حالت نرمال دوم قرار دارد یا خیر، باید آن را از نظر وابستگی تابعی بررسی کرد.

نمودار وابستگی تابعی مربوط به جدول SP را می‌توان به‌صورت زیر نشان داد:



شکل ۴-۲

همان‌طور که می‌دانید جدول SP از روی جداول S و P استخراج شده است، با دقت در جدول SP (جدول ۱۰-۲) می‌بینیم که اجزای کلید اصلی (S# و P#) به‌طور مجزا با Qty وابستگی تابعی ندارند و وابستگی تابعی فقط توسط ترکیبS# و P# وجود دارد، لذا می‌توان گفت که رابطه جدول SP نیز در حالت نرمال دوم۲ NF قرار دارد.

مثال: فرض کنید در محیط عملیاتی تولیدکنندگان قطعات جدول ۱۵-۲ را به جای SP در نظر گرفته‌ایم:

جدول ۱۵-۲

S#	Status	City	P#	Qty
S1	۲۰	تهران	P1	۳۰۰
S1	۲۰	تهران	P2	۲۰۰
S1	۲۰	تهران	P3	۴۰۰
S1	۲۰	تهران	P4	۲۰۰
S1	۲۰	تهران	P5	۱۰۰
S1	۲۰	تهران	P6	۱۰۰
S2	۱۰	کرمانشاه	P1	۳۰۰
S2	۱۰	کرمانشاه	P2	۴۰۰
S3	۱۰	کرمانشاه	P2	۲۰۰
S4	۲۰	تهران	P2	۲۰۰
S4	۲۰	تهران	P4	۳۰۰
S4	۲۰	تهران	P5	۴۰۰

این جدول در فرم نرمال اول قرار دارد اما در فرم نرمال نوع دوم قرار ندارد. اگر فیلدهای P#و S# را به عنوان کلید اصلی مرکب در نظر بگیریم، بخش‌هایی از کلید اصلی با سایر فیلدها وابستگی تابعی دارد.S# به تنهایی با City وابستگی تابعی دارد که این امر با تعریف فرم دوم نرمال در تناقض است و نباید اجزای کلید اصلی در یک رابطه با سایر فیلدها وابستگی تابعی داشته باشد.

۳-۴-۲ فرم سوم نرمال(۳NF)

موجودیت یا جدولی در فرم سوم نرمال است که اولاً در فرم دوم نرمال بوده، ثانیاً تمام صفت‌های غیرکلید آن وابستگی تابعی فقط به کلید اصلی داشته باشند نه به یک صفت غیرکلید، یا به عبارت دیگر صفات غیرکلید وابستگی تابعی با واسطه با کلید اصلی نداشته باشند به عنوان نمونه فرض کنید در محیط عملیاتی تولیدکنندگان قطعات، این جدول‌ها را در نظر گرفته‌ایم:

جدول ۲-۱۶

S C

S#	City
S1	تهران
S2	کرمانشاه
S3	کرمانشاه
S4	تهران
S5	بوشهر

جدول ۲-۱۷

C S

City	Status
بوشهر	۳۰
تهران	۲۰
کرمانشاه	۱۰
شیراز	۵۰

در این دو جدول می‌توان نمودار وابستگی تابعی را به‌صورت زیر ترسیم کرد:

شکل ۲-۵

به دلیل وجود واسطه در وابستگی تابعی، جدول‌های ۲-۱۶ و ۲-۱۷ در فرم سوم نرمال نیستند.

(در این جدول‌ها کلید S# با City وابستگی تابعی دارد و City نیز با Status وابستگی تابعی دارد).

برای حذف این وابستگی تابعی می‌توان فیلد واسطه را در وابستگی تابعی حذف کرد:

شکل ۲-۶

در این صورت وابستگی تابعی با واسطه حذف شده است، بنابراین رابطه در حالت سوم نرمال قرار دارد.

واژه‌نامه

صفت	Attribute
موجودیت	Entity
سلسله مراتبی	Hierarchical
شبکه	Network
نرمال‌سازی	Normalization
تولیدکننده	Producer
رابطه‌ای	Relational

خلاصه مطالب

- پایگاه داده مجموعه‌ای سازمان‌یافته از اطلاعات است که می‌تواند به‌صورت رکوردهای ذخیره شده در جداول باشد و با یک روش اصولی و از طریق فرم‌ها، گزارش‌ها و ... می‌تواند به درخواست‌های کاربران پاسخ دهد.
- مدل‌های پایگاه داده طرح و الگوی کاری برای کاربران پایگاه داده‌ها را در سطح منطقی مشخص می‌کند.
- مدل سلسله مراتبی توسط گراف پیاده‌سازی می‌شود که ارتباط بین گره‌ها از بالا به پایین صورت می‌گیرد و مسیری از گره‌های پایین‌تر به گره‌های بالاتر وجود ندارد.
- مدل شبکه‌ای شباهت زیادی به سلسله مراتبی دارد، اما به دلیل وجود ارتباط از پایین به بالا پیاده‌سازی آن راحت‌تر است.
- مدل رابطه‌ای یکی از بهترین و سریع‌ترین مدل‌های پایگاه داده است و اطلاعات در قالب جدول‌های مجزا مشخص می‌شوند.
- مدل رابطه‌ای بهترین مدل پیاده‌سازی سیستم‌های بانک‌های اطلاعاتی است. در این مدل پایگاه داده از دید کاربران شامل یک مجموعه جدول است که هر جدول دارای سطرها و ستون‌هایی است.
- تعیین ماهیت ارتباط در مدل رابطه‌ای یکی از مهم‌ترین مراحل طراحی بانک اطلاعاتی است و پس از تحلیل موجودیت‌ها و بررسی ماهیت ارتباط آن‌ها، باید رابطه‌ها را به‌صورت

جدول پیاده‌سازی کرد.

- پس از تهیه جدول‌های بانک اطلاعاتی به منظور بهینه‌سازی جداول، کاهش افزونگی اطلاعات و حفظ یکپارچگی اطلاعات باید آن‌ها را نرمال‌سازی کرد.
- نرمال‌سازی مدل‌های مختلفی دارد که مهم‌ترین آن‌ها عبارتند از:
- فرم نرمال اول
- فرم نرمال دوم
- فرم نرمال سوم