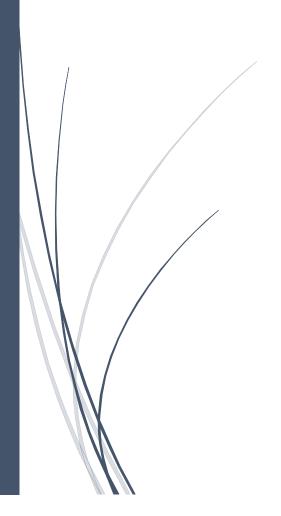
اصول طراحی بایگاه داده یا

حامد سپرزاده

دانشگاه فنی و حرفه ای، دانشگده شهید شمسی پور



مفهوم پایگاه داده از دهه ۱۹۶۰ برای کاهش مشکلات فزاینده در طراحی، ساخت، و نگهداشت سیستمهای اطلاعاتی معمولاً با تعداد زیادی کاربرِ نهایی همزمان، و با تعداد زیادی داده مختلف ایجاد شده است. اولین کاربردهای اصطلاح پایگاه داده به ژوئن ۱۹۶۳ باز می گردد، یعنی زمانی که شرکت System Development کاربردهای اصطلاح پایگاه داده به ژوئن ۱۹۶۳ باز می گردد، یعنی زمانی که شرکت Corporation مسئولیت اجرایی یک طرح به نام «توسعه و مدیریت محاسباتی یک پایگاه داده مرکزی» را بر عهده گرفت. پایگاه داده به عنوان یک واژه واحد در اوایل دهه ۷۰ در اروپا و در اواخر دهه ۷۰ در خبرنامههای معتبر آمریکایی به کار رفت.

اولین سیستم مدیریت پایگاه داده در دهه ۶۰ گسترش یافت. از پیشگامان این شاخه چارلز بکمن میباشد. مقالات بکمن این را نشان داد که فرضیات او کاربرد بسیار مؤثرتری برای دسترسی به وسایل ذخیره سازی را مهیا میکند. در آن زمان پردازش داده بر پایه کارتهای منگنه و نوارهای مغناطیسی بود که پردازش سری اطلاعات را مهیا میکرد. دو نوع مدل دادهای در آن زمان ایجاد شد CODASYL:که موجب توسعه مدل شبکهای شد که ریشه در نظریات بکمن داشت و مدل سلسله مراتبی که توسط North American شبکهای شد که ریشه در نظریات از آن شرکت IBM محصول کارا تولید نمود.

مدل رابطهای توسط E. F. Codd در سال ۱۹۷۰ ارائه شد. او مدلهای موجود را مورد انتقاد قرار می داد. برای در محتی نسبتاً طولانی این مدل در مجامع علمی مورد تأیید بود. در جریان سال ۱۹۸۰ پژوهش بر روی پایگاههای مدل توزیع شده و ماشینهای پایگاهی (database machines) متمرکز شد، اما تأثیر کمی بر بازار گذاشت. در سال ۱۹۹۰ توجهات به طرف مدل شیء گرا جلب شد. این مدل جهت کنترل دادههای مرکب لازم بود و بهسادگی بر روی پایگاه دادههای خاص، مهندسی داده(شامل مهندسی نرمافزار منابع) و دادههای چند رسانهای کار می کرد.

در سال ۲۰۰۰ نوآوری تازهای رخ داد و پایگاه اکسامال (XML) به وجود آمد. هدف این مدل از بین بردن تفاوت بین مستندات و دادهها است و کمک می کند که منابع اطلاعاتی چه ساخت یافته باشند یا غیرساخت یافته در کنار هم قرار گیرند.

فصل اول

بایگاه داده باچست ب

یک بانک اطلاعاتی یا پایگاه داده (database) مجموعه ای سازمان یافته و بدون افزونگی از دادههای مرتبط به هم است که می تواند توسط سیستمهای کاربردی مختلف به اشتراک گذاشته شود و به راحتی دسترسی، مدیریت و بهنگام شود.

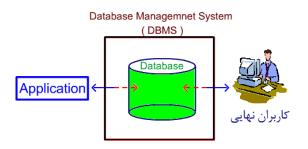
وقتی داده به صورت پایگاه داده سازماندهی میشود، کاربر و برنامه نویس نیازی به دانستن جزئیات ذخیره سازی داده ندارند. علاوه براین داده میتواند بدون تاثیر روی اجزای دیگر سیستم تغییر کند. برای مثال از اعداد حقیقی به صحیح، از یک ساختار فایل به دیگری یا از دستگاه ذخیره سازی نوری به مغناطیسی.

بنابراین می توان تعریف جامعی برای بانک اطلاعات ارائه کرد:

تُعْرِفُ بانک اطلاعات مجموعهای است از دادههای ذخیره شده به صورت مجتمع(یکپارچه) (نه لزوما فیزیکی، بلکه حداقل به طور منطقی)، بهم مرتبط، با کمترین افزونگی، تحت مدیریت یک سیستم کنترل متمرکز، مورد استفاده یک یا چند کاربر از یک یا بیش از یک "سیستم کاربردی"، به طور همزمان و اشتراکی.

سیستم مدیریت پایگاه دادهها (DBMS)

از سال ۱۹۵۰ ذخیره و بازیابی دادهها فراز و نشیب زیادی داشته است. در ابتدا کابران مستقیما با محیط سختافزار سر و کار داشتند تا اینکه نرمافزارهایی به عنوان رابط بین کاربر و سخت افزار ایجاد شدند. اما هنوز مشکلاتی وجود داشت. اشتراک دادهها، تامین امنیت و حفاظت از آنها مشکل بود. در اوایل دهه ۱۹۷۰ انقلابی در پایگاه دادهها بوجود آمد که به وجود آمدن حصار نفوذ ناپذیری به نام نظام مدیریت پایگاه دادهها (DBMS) بود. پایگاه دادهها تحت کنترل متمرکز نرمافزاری به نام سیستم مدیریت پایگاه دادهها ایجاد و بهره برداری می شود. می توان گفت که DBMS یک نرم افزار واسط ذخیره سازی و بازیابی اطلاعات است. هر کار کاربران به این سیستم تسلیم می شود و تا اگر این سیستم صلاح ببیند آن را انجام دهد یا رد کند. در واقع سیستم مدیریت پایگاه داده یک بسته نرم افزاری است که برای تعریف، بازیابی و مدیریت داده ها در یک پایگاه داده طراحی شده است. به طور کلی DBMS خود دادهها، قالب دادهها، نام فیلدها، ساختار رکوردها و ساختار فایل را مدیریت می کند. همچنین قوانینی برای اعتبار سنجی و مدیریت دادهها تعریف می کند.



سیستم مدیریت پایگاه دادهها (DBMS)

بند تعریف:

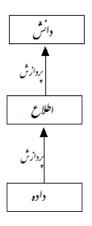
داده: مجموعه ای از کلمات یا اعداد بدون معنا یا با معنای ضمنی

¹ Data Base Management System

اطلاع: داده پردازش شده است (داده معنادار)

دانش: نوعی شناخت که از یک مجموعه اطلاعات حاصل میشود.

در شکل زیر ارتباط بین این سه مفهوم نمایش داده شده است. به عنوان مثال یک عدد مانند ۸۸۰۰۱۱ به تنهایی معنادار نیست و داده خام است. و اگر به ما توضیحی در مورد آن داده شود مثلا گفته شود که شماره دانشجویی است تبدیل به یک اطلاع می شود. این اطلاع با پردازش می تواند دانش هم حاصل کند. مثلا سال ورود دانشجو به دانشگاه ۸۸ است و این یک دانش است.



رابطه بین سه مفهوم داده، اطلاع و دانش

روشهای طراحی یک سیستم کاربردی

1) روش فایلی ۲) روش پایگاهی

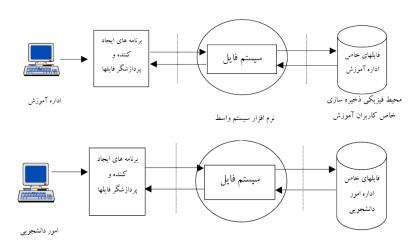
مشخصات روش فايلي

- √ تحلیل و بررسی نیازهای اطلاعاتی و پردازشی **هر قسمت به طور جداگانه**
- ✓ اجرای مراحل کلاسیک اولیه لازم برای طراحی و تولید یک سیستم کاربردی
 - ✓ تعیین مشخصات هر سیستم و وظایف آن
 - ✓ طراحی تعدادی فایل
 - ✓ نوشتن مجموعهای از برنامههای ایجاد، کنترل و پردازش فایل
 - ✓ استفاده از یک پیکربندی سختافزاری و نرمافزاری مشخص
 - ✓ انجام تستهای لازم و تنظیم سیستم کاربردی

√ ایجاد یک سیستم کاربردی برای هر قسمت و برپایی محیط فیزیکی ذخیره و بازیابی اطلاعات و سیستم بهرهبرداری از آن خاص همان قسمت.

مثالی از روش فایلی:

برای مثال سیستم دانشگاه را در نظر بگیرید که از سه بخش آموزش، اداره فارغ التحصیلان و اداره امور رفاهی دانشجویان تشکیل شده است. با توجه به روش فایلی این سه بخش از دانشگاه طراحی و پیاده سازی خود را از پایگاه دادهها در مورد دانشجو خواهند داشت.



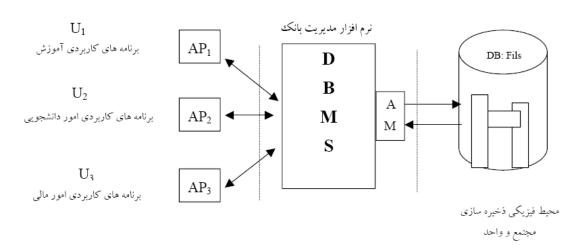
طراحی بانک اطلاعات دانشگاه به روش فایلی

معایب روش فایلی:

- ✓ عدم وجود محیط مجتمع ذخیرهسازی اطلاعات و عدم وجود سیستم یکپارچه
 - ✓ عدم وجود سیستم کنترل متمرکز روی کل دادهها
 - ✓ افزونگی: تکرار ذخیرهسازی در مورد دادهها
 - ✓ عدم وجود ضوابط ایمنی کارا و مطمئن
 - ✓ خطر بروز پدیده ناسازگاری دادهها
 - ✓ عدم امکان اشتراکی شدن دادهها
 - ✓ مصرف نابهینه امکانات سختافزاری و نرمافزاری
 - ✓ حجم زیاد برنامهسازی
 - ✓ وابستگی برنامههای کاربردی به محیط ذخیرهسازی دادهها

مشخصات روش پایگاهی

- ✓ بررسی و تحلیل نیازهای پردازشی و اطلاعاتی همه قسمتها توسط یک گروه
 - ✓ مدلسازی معنایی دادهها
 - ✓ تعیین مشخصات جامع (یکیارچه) کاربردی و وظایف آن
 - ✓ انتخاب یک یا چند پیکربندی سختافزاری-نرمافزاری
 - ✓ استفاده از یک یا چند DBMS
 - ✓ طراحی پایگاه دادهها در سطوح لازم
 - ✓ تولید مجموعهای از برنامههای ایجاد و کنترل پایگاه داده
 - ✓ ایجاد محیط واحد و مجتمع ذخیرهسازی و مشترک بین کاربران
 - ✓ طراحی و تولید واسطهای کاربرپسند مورد نیاز
 - ✓ تعریف پایگاه داده هر قسمت توسط کاربر مربوطه
 - ✓ طراحی برنامههای عملیات در پایگاه داده
 - ✓ بهرهبرداری واقعی از سیستم پس از تستهای لازم



طراحی بانک اطلاعات دنشگاه به روش پایگاهی

مشخصات روش پایگاهی

- ✓ دادههای مجتمع
- ✓ عدم وابستگی برنامههای کاربردی به دادهها و فایلها
 - ✓ تعدد شیوههای دستیابی به دادهها

- ✓ عدم وجود ناسازگاری در دادهها
 - ✔ اشتراکی بودن دادهها
 - ✓ امكان ترميم دادهها
 - ✓ کاهش افزونگی
 - 🗸 کاهش زمان تولید سیستمها
- ✓ امكان اعمال ضوابط دقيق ايمنى

عناصر محيط يايگاه دادهها

١) سخت افزار

- سخت افزار ذخیرهسازی دادهها (رسانه اصلی: دیسک با سرعت و ظرفیت بالا رسانه فرعی نـوار
 مغناطیسی برای تهیه نسخه پشتیبان و ثبت تراکنشها)
- نکته ۱: نسخه های پشتیبان و فایلهای ثبت تراکنشها، ابزارهای ترمیم، تامین ایمنی و کنترل جامعیت یایگاه داده ها هستند.
- نکته ۲: در مورد تراکنش در ادامه بحث می شود. در اینجا تنها دانستن اینکه یک تراکنش عملیاتی است که پایگاه داده را از یک حالت صحیح به حالت صحیح دیگری می برد کافی است.
 - سخت افزار پردازشگر (پردازنده)
- سخت افزار ارتباطی (رسانه های ارتباط بین کامپیوتر و دستگاه های جانبی: محلی برای یک
 سایت و شبکه ای برای پایگاه داده نامتمرکز)

نکته: معماری های پایگاه دادهها:

- 🖊 معماری متمرکز
- 🖊 معماری مشتری سرویس دهنده
 - 🖊 معماری توزیع شده
 - 🖊 معماری با پردازش موازی
 - 🖊 معماری چندپایگاهی
 - 🖊 معماری موبایل

۲) نرم افزار

- OBMS) سیستم مدیریت پایگاه دادهها
- Oracle کاربردی که در محیط DBMS اجرا می شوند. (Oracle DBC, Mysql)
 - نرم افزار شبکه

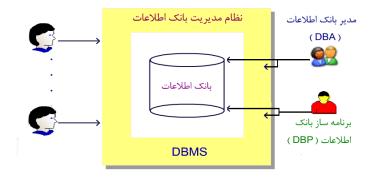
۳) کاربر

هر استفاده کننده از سیستم پایگاه دادهها

انواع كاربران:

- ✓ مدیر بانک اطلاعات (DBA): مسئولیت تصمیم گیری در مورد طراحی پایگاه دادهها را برعهده
 دارد. (یک نفر یا یک تیم)
 - ✓ برنامهساز بانک اطلاعات (DBP): تصمیمات مدیر را پیادهسازی می کند.
 - ✓ کاربر نهایی (End User): اپراتور یا استفاده کننده از پایگاه دادهها.

نکته: تنها کسانی که به دور از چشم DBMS می توانند به داده ها دسترسی داشته باشند، مـدیر و برنامه ساز یایگاه داده ها هستند.



دسترسی کاربران به بانک اطلاعات

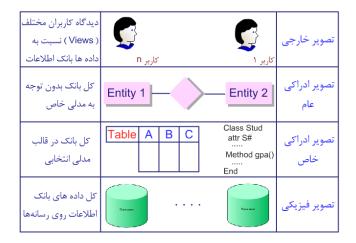
ها داده

دادههای ذخیره شده در پایگاه دادهها دادههای عملیاتی هستند. به این معنا که کاربران روزانه با آنها سرو کار دارند.

انواع دادهها: سیستمی و کاربران

معماری پایگاه دادهها

منظور از معماری پایگاه دادهها، چگونگی طراحی نرم افزار پردازش داده است که با توجه به نیاز به ایجاد رابط ه بین کاربران (در بالاترین سطح) و دادهها (در پایین ترین سطح) برپا میشود. یک مدل ۴ لایه ای توسط American National Standard Institute) ANSI ارائه شده است.



معماری چهار لایهای پایگاه دادهها

تصوير خارجي:

تنها لایه مربوط به کاربران ایمن لایه است (مانند: دانشجویان، مسئولین آموزش دانشکده)

در این لایه مواردی مانند تعریف سطح دسترسی به داده ها و تعریف دیدهای مختلف یا به عبارت دیگرتعریف دیدهایی ۲ مناسب با نیاز کاربران مطرح است.

قانون پیاده سازی داده ها می گوید: "به هر فردی به اندازه نیازش اطلاعات بدهید"



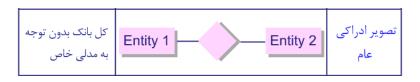
تصوير خارجي

تصویر ادراکی عام:

این تصویر به DBA و DBP مربوط می شود.

در مرحله طراحی بانک اطلاعات مطرح است و توسط برنامه نویسان به یک مدل قابل فهم برای کامپیوتر ترجمه میشود. بعد از طراحی نهایی و کامل بانک اطلاعات و انتخاب یک مدل برای پیادهسازی آن، این لایه به تصویر ادراکی خاص تبدیل میشود.

^{&#}x27; view



تصوير ادراكي عام

تصوير فيزيكى:

در این لایه ذخیره دادههای بانک اطلاعات بر روی رسانه مطرح است. این رسانهها به صورت مستقیم در اختیار کامپیوتر نیستند. آنچه مهم است، چگونگی ذخیرهسازی دادهها به همراه سرعت بالای خواندن و نوشتن است.

مهمترین موارد مورد بررسی در این لایه:

- پروندهها: شکلهای مختلفی که دادهها روی رسانه قرار می گیرند که سرعت بازیابی آنها سریع باشد.
 - شاخصها (index): ساختارهای اضافی برای دسترسی سریع به دادههای روی رسانه.



تصوير فيزيكى

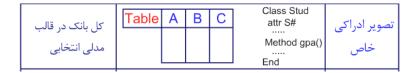
تصویر ادراکی خاص:

در این دید این موضوع مطرح است که چگونه می توان بین دیدهای مختلف کاربران و بانک اطلاعات ارتباط برقرار نمود؟

مفهوم انتزاع: بیان داده های پیچیده روی رسانه به شکل ساده تر دراختیار کاربران.

بهترین شکل: رابطه یا Relation(مفهوم تئوری)، جدول (حالت پیاده سازی)

مثال: جدولهای موجود در پایگاه داده یک دانشگاه.



تصویر ادراکی خاص

نکته: تصویر ادراکی عام تنها در مستندات وجود دارد و آنچه پیاده سازی شده و مورد استفاده قرار می گیرد تصویر ادارکی خاص است.

به ایجاد تصویر ادراکی پایگاه دادهها مدلسازی معنایی هم گفته میشود.

مدلهای یایگاه دادهها

۱) مدلهای قدیمی

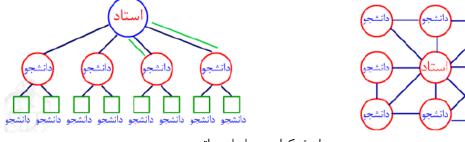
مدل سلسله مراتبی: از درخت برای سازماندهی دادهها استفاده می شود. (والد و فرزند)

در یک مدل سلسله مراتبی دادهها در یک ساختار درخت مانند سازماندهی شده که در آن هر رکورد دارای یک والد است. یک ستون مرتبسازی نیز رکوردهای خواهر/برادر را مرتب نگه میدارد. ساختار سلسله مراتبی در دوران سامانههای مدیریت بایگاه داده ابررایانهها (mainframe) از جمله در سامانههای مدیریت اطلاعات (IMS) به وسیله IBM به طور گسترده ای استفاده می شد و هماکنون الگویی از ساختار سندهای XML است.

مدل شبکهای: از گراف برای سازماندهی دادهها استفاده میشود.

مدل شبکهای تعمیم یافته ساختار سلسله مراتبی بوده و بنابراین امکان برقراری رابطه ٔ چند- به -چند را در ساختار درخت مانندی با امکان داشتن چندین والد را فراهم می کند. این مدل پیش از این که با مدل رابطه ای جایگزین شود پرطرفدارترین مدل داده بود. مدل شبکهای با بهره گیری از دو مفهوم دادهها را سازماندهی می کند که عبارتند از «رکوردها» و «مجموعهها .(sets) «رکوردها همانند زبان برنامهنویسی کوبول دارای ستونهایی هستند (که ممکن است به طور سلسله مراتبی سازماندهی شده باشند). مجموعهها (نباید با مجموعههای ریاضی اشتباه گرفته شوند) روابط یک - به - چند میان رکوردها تعریف می کنند: مانند یک مالک و بسیاری عضو. یک رکورد می تواند در هر تعداد مجموعه یک مالک و در بسیاری از مجموعهها عضو باشد.

هر دو مدل بر مبنای رکورد سازماندهی میشوند.



مدل شبکهای و سلسله مراتبی

مدل تخت: مدل تخت یا جدولی تشکیل شده است از یک آرایه دو بعدی با عناصر داده ای که همه اجزای یک ستون به صورت داده های مشابه فرض می شود و همه عناصر یک سطر با هم در ارتباط هستند. برای نمونه در ستون هایی که برای نام کاربری و رمز عبور در جزئی از سیستم های پایگاه داده ای امنیتی مورد استفاده قرار می گیرد هر سطر شامل رمز عبوری است که مخصوص یک کاربر خاص است.

۲) مدلهای رابطهای

بانک اطلاعات در مدل رابطه ای به صورت مجموعهای از رابطه ها طراحی می شود و دادهها به صورت رکوردهای مرتبط سازماندهی میشوند. عمده ترین دلیل موفقیت مدل رابطهای سادگی آن است و رابطه به صورت جدول پیادهسازی میشود. دلیل دیگر موفقیت مدل رابطهای پشتوانه تئوریک بسیار قوی آن است.

۳) مدلهای فرا رابطهای

پایگاه دادههای شامل صوت و تصویر و فیلم را به راحتی نمی توان در قلب رکورد اطلاعاتی درون رابطه ذخیره کرد. در سالهای اخیر مدلهایی پیشنهاد شدهاند مانند مدل منطقی، تابعی، شی گرا، شی – رابطهای، استنتاجی تا این مشکل را بر طرف نمایند.

چند تعریف

۱) دادهگان

به داده گان داده درباره داده هم گفته می شود. یعنی قصد داریم در مورد اطلاعات ذخیره شده در بانک اطلاعات اطلاعات دیگری ذخیره کنیم. داده گان در پایگاه داده ها به دو دسته تقسیم می شود.

لغتنامه دادهها

تمامی اسامی استفاده شده در پایگاه دادهها و معنای آنها را در برمیگیرد. مثلا دوجدول همنام نباید در یک پایگاه دادهها وجود داشته باشد.

كاتالوگ سيستم

هر سیستم مهمی دارای فایل کاتالوگ است که اطلاعات حقوق دسترسی افراد و اینکه چه کسی در چه زمانی چه تغییری را در پایگاه دادهها ایجاد کرده است را ذخیره می کند (فایل لاگ).

كاتالوگ سيستم اطلاعات لغتنامه دادهها را نيز در بر مي گيرد.

۲) امنیت

حفاظت دادهها در برابر خطرات طبیعی مانند آتشسوزی: تهیه نسخههای پشتیبان جلوگیری از دستیابی غیرمجاز دادهها: استفاده از رمز عبور و تغییر داده با الگوریتمهای خاص: هش کردن و رمزگذاری دادهها

[&]quot; Meta Data

جلوگیری از دستیابی غیرمجاز دادهها دو موضوع مطرح است.

احزار هویت (Authentication): کاربر همانی است که ادعا میکند.

سطح دسترسی (Authorization): کاربر چه دسترسی و حقوقی بر روی سیستم دارد.

۳) جامعیت

جامعیت به معنای $\frac{صحت}{}$ و $\frac{\text{سازگاری}}{}$ دادهها و پردازشهاست. برای مثال در پایگاه داده یک بانک موجودی افراد نمیتواند منفی باشد. یا اینکه فرد نمیتواند بیش از موجودی خود پول برداشت کند. جامعیت قواعدی دارد که در ادامه بررسی خواهند شد.

۴) تراکنش

هر برنامه ای که در محیط پایگاه داده ها توسط کاربر اجرا شود تراکنش نام دارد. تفاوت تراکنش با برنامه معمولی این است که تراکنش به سیستم مدیریت پایگاه داده ها تسلیم می شود. و DBMS می تواند آن را اجرا کرده، به تعویق انداخته یا ساقط کند. هدف از این گونه کنترل ها حفظ جامعیت و صحت پایگاه داده هاست.

خواص تراکنش ها در بانک اطلاعاتی (خواص ACID)

چهار کنترل زیر لازم است روی تمامی تراکنش ها در پایگاه دادهها اعمال گردد تا صحت و جامعیت آن تضمین شود.

(atomicity) یکیارچگی

یا همه دستورالعمل های یک تراکنش اجرا می شوند یا هیچ کدام. مثلا تراکنش انتقال پول از حسابی به حساب دیگر با کم شدن مبلغ از حساب اول و اضافه شدن به حساب مقصد انجام میشود و نباید نیمه کاره انجام شود.

✓ سازگاری (consistency)

هر تراکنش باید همه قوانین بانک اطلاعات را رعایت کند. مثلا در تراکنش انتقال پول از حسابی به حساب دیگر مبلغ برداشت شده از حساب اول با مبلغ واریز شده به حساب مقصد باید برابر باشد.

تراکنش می تواند دو نوع پایان داشته باشد:

موفق commit

ناموفق abort

✓ انزوا (Isolation)

اثرات تراکنشهای همروند روی یکدیگر به این شکل است که گویا هر کدام در انزوا انجام میشوند. تراکنشهای همروند نباید روی هم اثر مخرب بگذارند. واحد کنترل همروندی این عملیات را انجام میدهد.

✓ پایداری (Durability)

اثر تراکنشها ماندنی است و هرگز به طور اتفاقی از بین نمیرود. مثلا اگر به حسابی پولی ریخته شود با آتش سوزی در آن بانک آن مبلغ از بین نمیرود.

(۵) شمای پایگاه دادهها (Database Schema)

مجموعه ساختارهای طراحی شده در یک پایگاه دادهها بدون توجه به داده هایی که در آنها قرار می گیرند شمای پایگاه دادهها هستند. مثلا در پایگاه دادههای رابطه ای شما جدول ها و ستون ها و نوع دادههای هر ستون است ولی تعداد سطرها ربطی به شما ندارد چون هر لحظه ممکن است کم و زیاد شود.

(data dependency) استقلال دادهها (۶

استقلال داده ها به دو نوع تقسیم بندی می شود: فیزیکی و منطقی.

استقلال فیزیکی دادهها: نحوه ذخیره سازی اطلاعات روی رسانهها و تغییر در ذخیرهسازی دادهها برنامههای کاربردی تغییری نمی کنند.

استقلال منطقی دادهها: تغییر تصویر ادراکی بانک اطلاعات از دید کاربران و برنامه های کاربردی پنهان میماند. مثلا افزودن فیلدی به یک جدول پایگاه دادهها.

زبانهای برنامهسازی بانک اطلاعات

زبان پایگاه دادهها بیانی است یعنی فقط در آن گفته می شود چه میخواهیم و طریقه انجام کار را مانند زبانهای دستوری بیان نمی کنیم. در پایگاه دادهها دو زبان وجود دارد:

- نعریف دادهها (DDL): تعریف و تغییر ساختار پایگاه دادهها
 - ✓ زبان کار با دادهها (DML): دستیابی، تغییر و یا بازیابی دادهها

BASIS FOR COMPARISON	DDL	DML
Basic	DDL is used to create the database schema.	DML is used to populate and manipulate database
Full Form	Data Definition Language	Data Manipulation Language
Classification	DDL is not classified further.	DML is further classified as Procedural and Non- Procedural DMLs.
Commands	CREATE, ALTER, DROP, TRUNCATE AND COMMENT and RENAME, etc.	SELECT, INSERT, UPDATE, DELETE, MERGE, CALL, etc.

به طور خلاصه مزایا و ویژگی های پایگاه دادهها:

- ۱) اشتراک دادهها (shared) داده ها در پایگاه داده بین کاربران و برنامه های کاربردی به اشتراک گذاشته می شوند.
- ۲) ماندگاری(persistence)
 وقتی داده در پایگاه داده ذخیره شد پایدار است و از بین نمی رود، مگر آنکه توسط سیستم پایگاه داده
 تغییر کند.
 - ۳) امنیت(security)

داده در پایگاه داده از فاش شدن، تغییر و تخریب بدون مجوز محافظت می شود. مدیر سیستم توسط سطوح دسترسی و قیدهای امنیتی نحوه دستیابی به داده را تعریف می کند و اطمینان می دهد که دستیابی از طریق مناسب انجام می شود.

- (validity) و صحت (۴) اعتبار
 - (integrity) جامعیت (۵
- ۶) سازگاری(consistency)
 داده در پایگاه داده با مقدار واقعی داده در دنیای خارج سازگار است .وقتی یک فقره اطلاع در بیش از

- یک نقطه ذخیره شود و لازم باشد بهنگام شود، اگر بهنگام سازی در همه نقاط انجام نشود ناسازگاری ایجاد می شود.
- ۷) کاهش افزونگی (non redundancy) داده در پایگاه داده دارای حداقل افزونگی است. تکرارهای بی مورد نباید وجود داشته باشد که منجر به استفاده از فضای بی مورد شود.
 - ۸) سهولت دسترسی به دادهها
 - ۹) شاخص گذاری و سایر امکانات ذخیره سازی کارا
 - ۱۰) ذخیره و بازیابی دادهها تحت یک کنترل متمرکز
 - ۱۱)امکان ترمیم (Recovery) دادهها در صورت وقوع خطا، سیستم به حالت صحیح قبلی بازمی گردد.
 - (independence)استقلال(۱۲
- تغییر در نمایش فیزیکی، تکنیک های دستیابی و سازماندهی داده تاثیری روی برنامه های کاربردی ندارد. استقلال فیزیکی و منطقی
 - ۱۳)دستیابی همروند به دادهها

. فصل دوم

مدل سازی معنایی داده کو

همانطور که گفتیم به طراحی ادراکی پایگاه داده ها مدل سازی معنایی داده ها گویند. روش رایج کلاسیک برای مدل سازی معنایی داده ها روش موجودیت - ارتباط † یا ER است. در این روش سه مفهوم معنایی وجود دارد: نوع موجودیت، صفت، نوع ارتباط.

نوع موجودیت: هر چیزی که میخواهیم در مورد آن اطلاع داشته باشیم.

مثال موجودیتهای سیستم انتخاب واحد دانشگاه را نام ببرید:

دانشجو — درس — استاد — گروه درسی

مثال موجودیتهای یک سیستم صدور بلیط اتوبوسرانی را نام ببرید:

اتوبوس — راننده — سفر — شهر – مسافر

مثال موجودیتهای یک سیستم مطب را نام ببرید:

بيمار – نسخه – دارو

موجودیت را با یک مستطیل در نمودار ER نمایش می دهیم.

^{*} Entity-Relation

نمونه موجودیت: یک نوع موجودیت چندین نمونه موجودیت دارد که از هم متمایز هستند. مثلا نوع موجودیت درس می تواند دارای دو نمونه موجودیت پایگاه دادهها و کامپایلر باشد.

نکته: هر چیزی معمولا موجودیت است، که نمیدانیم چند نمونه دارد.

صفت: هر نوع موجودیت چند صفت دارد که ویژگیهای آن موجودیت را بیان می کنند.

برای مثال در سیستم ثبتنام دانشگاه صفات هر موجودیت به صورت زیر است:

دانشجو: شماره دانشجویی، نام، نام خانوادگی، شماره تلفن، آدرس

درس: کد درس، نام درس، تعداد واحد، نوع درس

استاد: کد استاد، نام، نام خانوادگی، مدرک، رشته

گروه درسی: کدگروه، شماره گروه

صفات را با بیضی در نمودار ER نمایش میدهیم.

انواع صفات:

۱) ساده یا مرکب

صفت ساده: صفتی است که به جزء معناداری تجزیه پذیر نباشد. مثلا نام دانشجو ساده است. صفت مرکب: صفتی است که از چند صفت ساده تشکیل می شود. مثلا آدرس مرکب است چون از چند صفت ساده شهر، خیابان، کوچه و پلاک تشکیل می شود.

۲) شناسه (کلید)

صفت کلیدی صفت یا مجموعه صفاتی است که در یک موجودیت یکتا و منحصر به فرد باشد. مثلا در موجودیت دانشجو شماره دانشجویی، در موجودیت درس کد درس و در موجودیت استاد کد استاد در یک دانشگاه منحصر به فردند و صفت کلیدی هستند. صفت کلیدی را با یک خط درون بیضی مشخص می کنند.

۳) تک مقداری یا چندمقداری

صفت تک مقداری تنها یک مقدار دارد و صفت چند مقداری می تواند چند مقدار بگیرد. مثلا هر فردی یک نام و یک نام خانوادگی دارد اما می تواند چندین مدرک داشته باشد. بنابراین مدرک می تواند یک صفت چند مقداری باشد. یا شماره تلفن های یک شرکت می تواند متعدد باشد.

صفت چند مقداری با دو خط به موجودیت متصل می شود.

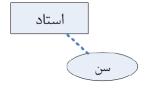


۴) مشتق یا غیر مشتق

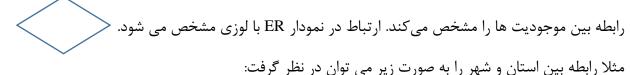
صفت مشتق صفتی است که به عنوان صفتی از یک موجودیت وجود خارجی ندارد اما از روی سایر صفات قابل حصول است. مثلا:

- معدل دانشجو از روی نمرات او قابل محاسبه است. بنابراین معدل دانشجویان بهتر است صفت مشتق باشد. اما معدل کل دانشجویان فارغ التحصیل شده بهتر است به عنوان صفتی از موجودیت در نظر گرفته شود.
- سال ورود دانشجو، دانشکده، رشته و نوبت تحصیلی (روزانه، شبانه) از روی شماره دانشجویی او
 قابل دریافت است.
 - شهر افراد از روی شماره ملی آن ها قابل دستیابی است.
 - سن افراد از تاریخ تولد آن ها قابل محاسبه است.

صفت مشتق با نقطه چین به موجودیت متصل می شود.



ار تباط:



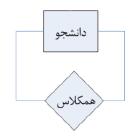


رابطه بین استاد و درس:

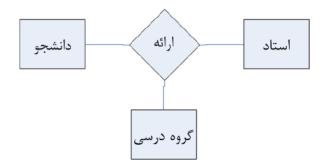


تقسیم بندی ارتباطات:

از نظر درجه ارتباط : به تعداد موجودیت هایی که در ارتباط شرکت می کنند درجه ارتباط گویند. مثلا هر دو نمودار بالا در جه ۲ هستند. نمودار ER زیر درجه یک است:



یک ارتباط درجه ۳:



ارتباطات بیش از درجه ۳ کمیاب هستند.

اتصال

ارتباطها بر اساس نوع ارتباط به سه نوع تقسیم می شوند.

۱-۱ (یک به یک)

1-N (یک به چند)

N-M (چند به چند)

ارتباط یک به یک به این معناست که یک نمونه موجودیت از موجودیت اول تنها با یک نمونه موجودیت از موجودیت دوم می تواند رابطه داشته باشد.

مثلا رابطه اداره و مدیرکل را در نظر بگیرید. یک اداره تنها یک مدیرکل دارد. (رابطه ۱-۱)

ارتباط یک به چند به این معناست که یک نمونه موجودیت از موجودیت اول می تواند با یک یا چند نمونه موجودیت از موجودیت دوم می تواند با یک نمونه موجودیت از موجودیت دوم می تواند با یک نمونه موجودیت از موجودیت اول در ارتباط باشد.

مثلا رابطه اداره و کارمند را در نظر بگیرید. یک اداره چند کارمند دارد. و هر کارمند تنها در یک اداره کار میکند (رابطه N-1)



ارتباط چند به چند به این معناست که یک نمونه موجودیت از موجودیت اول می تواند با یک یا چند نمونه موجودیت از موجودیت دوم می تواند با چند نمونه موجودیت از موجودیت دوم می تواند با چند نمونه موجودیت از موجودیت اول در ارتباط باشد.

مثلا رابط استاد و درس را در نظر بگیرید. یک استاد می تواند چند درس را ارائه کند و هر درس می تواند توسط چند استاد ارائه شود. یا یک استاد می تواند چند دانشجو داشته باشد و هر دانشجو هم می تواند چند استاد داشته باشد.



حد ارتباط:

حد ارتباط حداقل و حداکثر نمونه موجودیت هایی از ارتباط را مشخص میکند که در ارتباط شرکت می کنند. در مثال زیر یک استاد می تواند ۰ تا ۱۰ دانشجو را راهنمایی کند و یک دانشجو ممکن است استاد راهنما نداشته باشد (۰) و یا یک استاد راهنما داشته باشد.



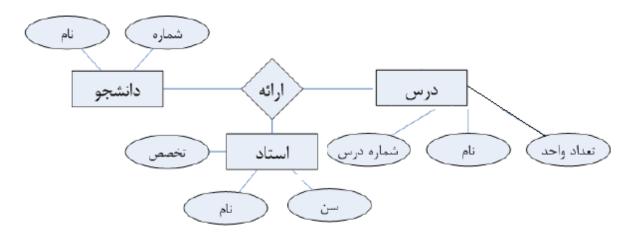
شرکت اختیاری یا اجباری در ارتباط

در نمودار ER زیر شرکت دانشجو در ارتباط اختیاری است یعنی دانشجو می تواند استاد راهنما نداشته باشد. شرکت اختیاری در ارتباط با یک دایره توخالی در طرف مقابل موجودیت نمایش داده میشود.



یک نمودار ER کامل تر:

نمودار ER ثبت نام دانشگاه را رسم کنید:

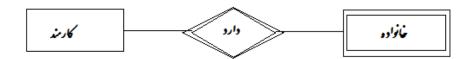


به نظر شما جایگاه صفات نیمسال تحصیلی و نمره کجاست؟

موجوديت ضعيف

موجودیتی است که وجودش وابسته به موجودیت قوی دیگری است و با حذف موجودیت قوی از سیستم موجودت ضعیف هم حذف می شود.

مثلا خانواده کارمند برای کارمند موجودیت ضعیف است. موجودیت ضعیف با دو مستطیل تودرتو و رابطه ضعیف (بین یک موجودیت ضعیف و قوی) با لوزی تودرتو نشان داده می شود.



انواع كليد

ابرکلید (سوپر کلید): هر کلید یا مجموعه ای از کلیدها که خاصیت یکتایی مقدار دارند. مثلا در موجودیت دانشجو ترکیب همه دانشجو ترکیب شماره دانشجویی و نام دانشجو ابرکلید است چون یکتاست. در همین موجودیت ترکیب همه صفات هم ابر کلید است. شماره دانشجویی هم ابر کلید است چون یکتاست. بنابراین هر چه با شماره دانشجویی ترکیب شود نتیجه ابر کلید است که یکتاست.

کلید کاندید: ابر کلیدی است که کمینه باشد. برای مثال در موجودیت دانشجو ترکیب شماره دانشجویی و نام دانشجو ابر کلید است اما کلید کاندید نیست چون کمینه نیست. و این به این دلیل است که جزء کوچکتری از آن به نام شماره دانشجویی خود خاصیت یکتایی مقدار دارد. پس شماره دانشجویی یک کلید کاندید می شود چون هم یکتا و هم کمینه است.

کلید اصلی: یکی از کلیدهای کاندید به انتخاب طراح بانک اطلاعات به عنوان کلید اصلی در نظر گرفته میشود.

کلید بدیل (جایگزین یا فرعی): کلیدهای کاندید بجز کلید اصلی، کلیدهای جایگزین کلید اصلی هستند.

کلید خارجی: کلیدی است که در یک رابطه کلید کاندید است و در رابطه دیگر صفتی ساده و برای ارتباط وارد شده است.

قواعد جامعيت

قواعد جامعیت به سه قاعده تقسیم بندی می شوند:

درون رابطهای: این قاعده جامعیت می گوید هر رابطه به خودی خود باید صحیح باشد. به این معنا که کلید اصلی (کاندید) آن شناخته شده باشد. کلید اصلی (کاندید) نباید

ارجاعی: ناظر به کلید خارجی است. طبق این قاعده کلید خارجی مقدار خود را باید از کلید کاندید خود بگیرد. و مقداری خارج از مقادیر کلید کاندید نداشته باشد.

دامنهای: هر صفت یا فیلد باید مقادیرش را از دامنه خود بگیرد. مثلا شماره دانشجویی نمی تواند حرف داشته باشد یا نام فرد نمی تواند عدد داشته باشد.

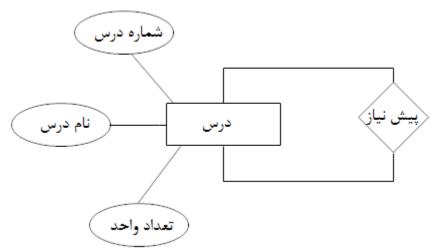
تمرین: برای سیستمهای ذکر شده زیر نمودار ER طراحی نمایید:

- ✓ مطب (دارای یک پزشک است که اطلاعات بیماران مراجعه کننده را و اینکه چه بیماری داشته و چه نسخه ای برای آن ها تجویز شده است را نگهداری می کند)
 - ✔ بانک (اطلاعات مشتریان و حساب های آن ها و مبادات اینترنتی، نقدی، پوز و عابر بانک)
 - ✓ بنگاه معاملات ملکی
 - ✓ فروشگاه کتاب
 - ✓ کتابخانه با قابلیت رزرو کتاب
 - ✓ شرکت مسافربری اتوبوس
 - ✓ آزانس مسافربری
- ✓ آموزشگاه (آموزشگاه دوره هایی را با ظرفیت خاصی اعلام میکند، متقاضیان در دوره ثبت نام میکنند و اگر تعداد متقاضیان به حد مشخص شده رسید مبلغ لازم را پرداخت میکنند و دوره برگزار میشود)

تبدیل نمودار ER به جدول

۱) ارتباط N-M درجه یک

برای تنها موجودیت، یک جدول در نظر می گیریم و جدول دومی را برای ارتباط ایجاد می کنیم و کلید اصلی جدول موجودیت را به عنوان کلید خارجی دو بار در جدول ارتباط وارد می کنیم: برای مثال ارتباط زیر را در نظر بگیرید:



این ارتباط یک ارتباط چند به چند است. یعنی هر درس می تواند پیش نیاز چند درس باشد و هر درس می تواند چند پیش نیاز داشته باشد.

تبدیل به جدول:

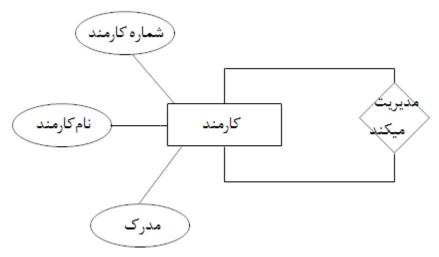
درس(شماره درس، نام درس، تعداد واحد) پیشنیاز (شماره درس درس ۱، شماره درس ۲)

پیش نیاز				
پیش	شماره درس	شماره درس		
	نياز			
	1111	1117		
	1111	1117		
	1111	1111		

درس				
تعداد واحد	نام درس	شماره درس		
٣	پایگاه داده ها	1111		
٣	مهندسی نرم افزار	1117		
٣	مدار منطقی	1117		
٣	معماری کامپیوتر	1114		

۲) ارتباط I-N درجه یک

مانند ارتباط N-M درجه یک، تبدیل به جدول می شود.

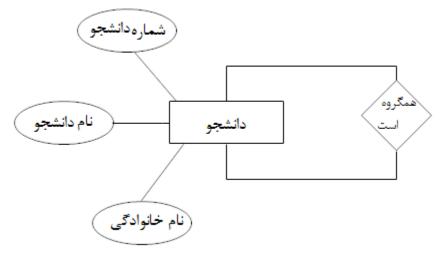


این ارتباط یک به چند است. یعنی یک کارمند می تواند مدیر چند کارمند باشد و هر کارمند هم یک مدیر دارد. (فرض کنید مدیریت بخشی است که هر بخش چند کارمند دارد که یکی از آن ها مدیر سایرین است.) کارمند (کد کارمند، نام کارکند، نام خانوادگی، مدرک، آدرس، تلفن) مدیریت (کد کارمند مدیر، کد کارمند تحت مدیریت) ترکیب دو صفت کلید اصلی است.

٣) ارتباط 1-1 درجه يک

برای ارتباطات 1-1 درجه یک، تنها یک جدول برای موجودیت ایجاد می شود و کلید اصلی جدول به عنوان کلید خارجی وارد خود جدول می شود.

فرض کنید در انجام پایان نامه دانشجویی قرار است پایان نامه ها دو نفره انجام شود یعنی هر دانشجویی یک هم گروهی دارد:



دانشجو (شماره دانشجویی، نام، نام خانوادگی، آدرس، تلفن، شماره دانشجوی همگروه)

دانشجو					
شماره دانشجوی هم گروه	آدرس	نام خانوادگی	نام	شماره دانشجو	
۸۸۰۰۲	تهران	حسيني	محمد	۸۸۰۰۱	
۸۸۰۰۱	تهران	کرمی	رضا	۸۸۰۰۲	
۸۸۰۰۴	تهران	نادري	زهرا	۸۸۰۰۳	
۸۸۰۰۳	تهران	غفاري	على	۸۸۰۰۴	

برای این ارتباط هم می توان جدول دومی ایجاد کرد: گروه (شماره دانشجوی ۱ ، شماره دانشجوی ۲)

گروه	
شماره دانشجوی هم گروه	شماره دانشجو

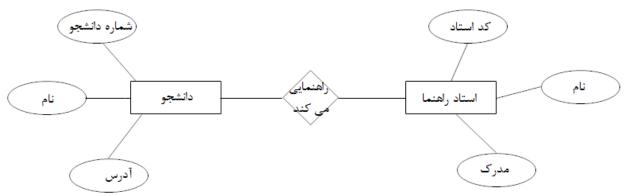
۸۸۰۰۲	۸۸۰۰۱
۸۸۰۰۴	۸۸۰۰۳

دانشجو						
آدرس	نام خانوادگی	نام	شماره			
			دانشجو			
تهران	حسيني	محمد	۸۸۰۰۱			
تهران	کرمی	رضا	۸۸۰۰۲			
تهران	نادري	زهرا	۸۸۰۰۳			
تهران	غفاري	على	۸۸۰۰۴			

سطرهای تکراری در جدول گروه حذف شد اند.

۴) ارتباط 1-N درجه دو

برای هر یک از موجودیت ها یک جدول ایجاد می شود و کلید اصلی موجودیت طرف یک به عنوان کلید خارجی وارد جدول موجودیت طرف N می شود.



این ارتباط یک به چند است. یعنی یک استاد می تواند چند دانشجو را در انجام پایان نامه راهنمایی کند و هر دانشجو می تواند یک استاد راهنما داشته باشد.

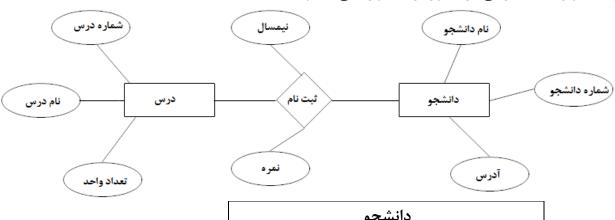
استاد					
مدرک	نام خانوادگی	نام	کد استاد		
دکتری	رضایی	محمود	111		
ارشد	نعيمى	محسن	177		
ارشد	مرادي	ميلاد	١٣٣		

دکتری	نوری	فاطمه	144
-------	------	-------	-----

دانشجو					
کد استاد راهنما	آدرس	نام خانوادگی	نام	شماره	
				دانشجو	
111	تهران	حسيني	محمد	۸۸۰۰۱	
111	تهران	کرمی	رضا	۸۸۰۰۲	
177	تهران	نادري	زهرا	۸۸۰۰۳	
177	تهران	غفاري	على	۸۸۰۰۴	

۵) ارتباط N-M درجه دو

برای هر یک از موجودیت ها یک جدول در نظر می گیریم و یک جدول هم برای ارتباط ایجاد می کنیم و صفات ارتباط را در صورت وجود در آن وارد میکنیم. سپس کلید اصلی دو جدول موجودیت را به عنوان کلید خارجی در جدول ارتباط وارد می کنیم.



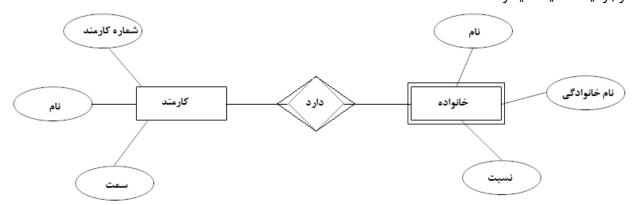
دانشجو						
آدرس	نام خانوادگی	نام	شماره			
			دانشجو			
تهران	حسيني	محمد	۸۸۰۰۱			
تهران	کرمی	رضا	۸۸۰۰۲			
تهران	نادری	زهرا	۸۸۰۰۳			
تهران	غفارى	على	۸۸۰۰۴			

درس

ا نام درس	تعداد واحد					
	٣	ثبت نام				
مهندسی نرم افزار ۳	٣	شماره دانشجو	شماره درس	نيمسال	نمره	
مدار منطقی ۳	٣	۸۸۰۰۱	1111	971	19	
معماری کامپیوتر ۳	٣	۸۸۰۰۱	1114	971	۲٠	
		۸۸۰۰۲	1117	971	١٧	
		۸۸۰۰۲	1114	971	١٨	
جوديت ضعيف		۸۸۰۰۳	1111	971	۲٠	

۶) ارتباط دارای موجودیت

برای هر موجودیت یک جدول و کلید اصلی موجودیت قوی به عنوان کلید خارجی وارد جدول موجودیت ضعیف میشود.



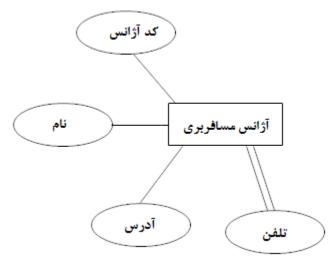
كارمند				
سمت	نام خانوادگی	نام	کد کارمند	
معاون	رضایی	محمود	111	
مدير بخش	نعيمى	محسن	177	
كارمند	مرادي	ميلاد	188	
كارمند	نوری	محمد	144	

خانواده				
کد کارمند	نسبت	نام خانوادگی	نام	کد
111	همسر	کریمی	فاطمه	١
111	فرزند	رضایی	على	٢

111	فرزند	رضایی	زهرا	٣
177	همسر	ناصري	ناهید	*
١٣٣	همسر	بابلى	مريم	۵
144	همسر	فرزادي	مينا	۶
144	فرزند	نوری	منا	٧

۷) موجودیت دارای صفت چند مقداری

برای موجودیت یک جدول و برای صفت چند مقداری هم یک جدول ایجاد می کنیم و کلید اصلی جدول موجودیت به عنوان کلید خارجی وارد جدول صفت چند مقداری می شود.

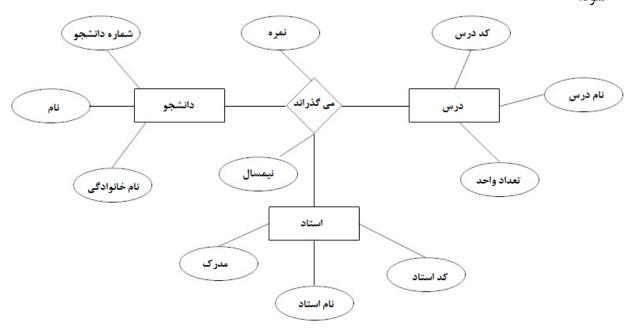


آژانس				
آدرس	نام	کد		
تهران	یاس	11		
تهران	گلبرگ	77		
تهران	فريد	٣٣		
تهران	نسترن	44		

آژانس		
تلفن	کد	
۸۷۸۷	11	
۲۲۸۷	11	
9,49	77	
٨٨٨٨	77	
۸۸۸۷	77	
7666	٣٣	

۸) ارتباط با درجه بیشتر از ۲

برای هر یک از موجودیت ها یک جدول و برای ارتباط هم یک جدول ارتباط می جدول ایجاد میشود. کلید اصلی همه موجودیت ها به عنوان کلید خارجی وارد جدول ارتباط می شود.



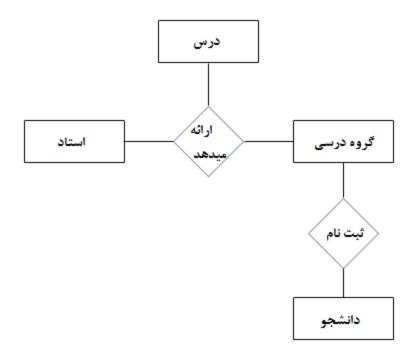
دانشجو				
آدرس	نام خانوادگی	شماره		
			دانشجو	
تهران	حسيني	محمد	۸۸۰۰۱	
تهران	کرمی	رضا	۸۸۰۰۲	
تهران	نادري	زهرا	۸۸۰۰۳	
تهران	غفارى	على	۸۸۰۰۴	

استاد				
مدرک	نام خانوادگی	نام	کد استاد	
دکتری	رضایی	محمود	111	
ارشد	نعيمى	محسن	177	
ارشد	مرادي	ميلاد	177	
دکتری	نوری	فاطمه	144	

درس				
تعداد واحد	نام درس	شماره درس		
٣	پایگاه داده ها	1111		
٣	مهندسی نرم افزار	1117		
٣	مدار منطقی	1117		
٣	معماری کامپیوتر	1114		

	ثبت نام			
نمره	نيمسال	کد استاد	شماره درس	شماره دانشجو
19	971	111	1111	۸۸۰۰۱
۲٠	971	١٢٢	1114	۸۸۰۰۱
١٧	971	١٣٣	1111	۸۸۰۰۲
١٨	971	١٢٢	1114	۸۸۰۰۲
۲٠	971	111	1111	۸۸۰۰۳
١٩	971	111	1111	۸۸۰۰۲
١٩	971	١٣٣	1111	۸۸۰۰۲

تمرین نمودار ER زیر را به جدول تبدیل کنید.



تمرین: نمودار ER طراحی شده برای سیستم های زیر را به جدول تبدیل کنید:

- ✓ مطب (دارای یک پزشک است که اطلاعات بیماران مراجعه کننده را و اینکه چه بیماری داشته و چه
 نسخه ای برای آن ها تجویز شده است را نگهداری می کند)
 - ✓ بانک (اطلاعات مشتریان و حساب های آن ها و مبادات اینترنتی، نقدی، پوز و عابر بانک)
 - ✓ بنگاه معاملات ملکی
 - ✓ فروشگاه کتاب
 - ✓ کتابخانه با قابلیت رزرو کتاب
 - ✓ شرکت مسافربری اتوبوس
 - 🗸 آزانس مسافربری
- ✓ آموزشگاه (آموزشگاه دوره هایی را با ظرفیت خاصی اعلام میکند، متقاضیان در دوره ثبت نام میکنند و اگر تعداد متقاضیان به حد مشخص شده رسید مبلغ لازم را پرداخت میکنند و دوره برگزار میشود)