

فهرست

اهداف این درس	۲
مفاهیم پایگاه داده‌ها	۲
سیستم ذخیره و بازیابی اطلاعات در معنای عام	۲
رده‌های تکنولوژیکی سیستم مدیریت پایگاه داده‌ها	۲
داده	۴
تعریف داده از دیدگاه ANSI	۴
تعریف اطلاع	۴
تعریف دانش	۴
تعریف پایگاه داده‌ها	۴
مراحل کلی کار در مشی فایلینگ	۴
معایب مشی فایلینگ	۵
مراحل کلی کار در مشی پایگاهی	۶
عناصر محیط پایگاه داده‌ها	۷
انواع سخت‌افزارهای محیط پایگاه داده	۷
انواع نرم‌افزارهای موجود در محیط پایگاه داده‌ها	۷
مدلسازی معنایی داده‌ها	۷
انواع روشهای مدلسازی معنایی داده‌ها	۸
سه مفهوم معنایی موجود در روش ER	۸
تعریف موجودیت	۸
سه ضابطه در رابطه با تشخیص يك نوع موجودیت	۸
موجودیت مستقل و وابسته	۸
تعریف صفت	۸
انواع صفت	۸
خصوصیات نوع ارتباط	۹

۹	نمودار ER
۱۴	مدل رابطه‌ای
۱۴	تعریف رابطه
۱۵	تناظر بین مفاهیم رابطه‌ای و مفاهیم جدولی
۱۵	ویژگی‌های رابطه
۱۵	انواع رابطه
۱۵	میدان (دامنه)
۱۵	مزایای میدان
۱۶	رابطه نرمال و غیر نرمال
۱۶	دلیل لزوم نرمال بودن رابطه
۱۶	معایب رابطه نرمال
۱۶	مزایا و معایب رابطه غیرنرمال
۱۷	کلید در مدل رابطه‌ای

اهداف این درس

در این درس به بخشی از مفاهیم مبنایی دانش و تکنولوژی پایگاه داده‌ها و اصول مدلسازی و طراحی آن پرداخته می‌شود. پایگاه داده و عناصر اصلی محیط آن معرفی می‌شود. با مدلسازی معنایی داده‌ها و محیط انتزاعی آشنا می‌شویم. سطوح معماریهای پایگاه داده‌ها بیان می‌شود. سیستم مدیریت پایگاه داده‌ها (DBMS) و اجزاء آن شرح داده می‌شود. به انواع معماریهای سیستم پایگاهی و مفاهیم اساسی مدل رابطه‌ای پرداخته می‌شود. با زبان SQL به عنوان یک زبان رابطه‌ای و همچنین با دیدهای رابطه‌ای آشنا می‌شویم و در نهایت طراحی پایگاه داده‌ها به روش بالا به پایین و طراحی فیزیکی را خواهیم دید.

مفاهیم پایگاه داده‌ها

اصطلاح پایگاه داده‌ها یکی از رایج‌ترین اصطلاحات در دانش و فن کامپیوتر است. سیستم مدیریت پایگاه داده‌ها یکی از سیستم‌های ذخیره و بازیابی اطلاعات است.

۱- سیستم ذخیره و بازیابی اطلاعات در معنای عام:

هر سیستمی که به کاربر برنامه‌ساز یا نابرنامه‌ساز امکان دهد تا اطلاعات خود را ذخیره، بازیابی و پردازش کند.

۲- رده‌های تکنولوژیکی سیستم مدیریت پایگاه داده‌ها

- ۱- سیستم فایلینگ
- ۲- سیستم مدیریت داده‌ها
- ۳- سیستم مدیریت پایگاه داده‌ها
- ۴- سیستم مدیریت پایگاه شناخت
- ۵- سیستم مدیریت پایگاه داده‌های شیء‌گرا
- ۶- سیستم هوشمند مدیریت پایگاه داده‌ها
- ۷- سیستم معنایی مدیریت پایگاه داده‌ها
- ۸- سیستم مدیریت پایگاه داده‌های زمانبند
- ۹- سیستم مدیریت پایگاه داده‌های نیم‌ساختمند و ناساختمند
- ۱۰- سیستم مدیریت پایگاه داده‌های بی‌درنگ
- ۱۱- سیستم داده‌کاوی و کشف شناخت
- ۱۲- سیستم مدیریت چند پایگاهی
- ۱۳- سیستم اطلاعات اجرایی
- ۱۴- سیستم فعال مدیریت پایگاه داده‌ها
- ۱۵- سیستم مدیریت پایگاه داده‌های شیء-رابطه‌ای

۳- داده

تعریف اول- نمایش ذخیره شده اشیاء فیزیکی، چیزهای مجرد، پدیده‌ها، رویدادها یا چیزهای قابل مشاهده که در تصمیم‌سازی به کار می‌آیند.

تعریف دوم- هر مجموعه‌ای از پدیده‌ها

تعریف سوم- پدیده‌های خام که معنای اندکی دارند مگر اینکه به صورت منطقی سازمان‌دهی شده باشند

تعریف داده از دیدگاه ANSI

نمایش واقعیتها، پدیده‌ها، مفاهیم یا شناخته‌ها به طور صوری و مناسب برای برقراری ارتباط، تفسیر یا پردازش توسط انسان یا هر امکان خودکار

هر نمایشی اعم از کاراکتری یا کمیتهای قیاسی که معنایی به آن قابل انتساب باشد.

۴- تعریف اطلاع

اطلاع به داده‌ای اطلاق می‌شود که توسط یک فرد یا سازمان برای تصمیم‌گیری بکار می‌رود
اطلاع، داده پردازش شده است.

اطلاع عبارت است از داده سازمان یافته‌ای که شناختی را منتقل می‌کند

۵- تعریف دانش

دانش عبارتست از نمایش نمادین جنبه‌هایی از بخشی از خرد جهان واقع

۶- تعریف پایگاه داده‌ها

مجموعه‌ای است از داده‌های ذخیره شده و پایا، به صورت مجتمع (یکپارچه) (نه لزوماً 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

برای ایجاد یک سیستم کاربردی دو راهکار وجود دارد:

۱- راهکار سنتی یا مشی فایلینگ

۲- راهکار (مشی) پایگاهی

۷- مراحل کلی کار در مشی فایلینگ

- تحلیل و بررسی نیازهای اطلاعاتی و پردازشی هر قسمت به طور جداگانه

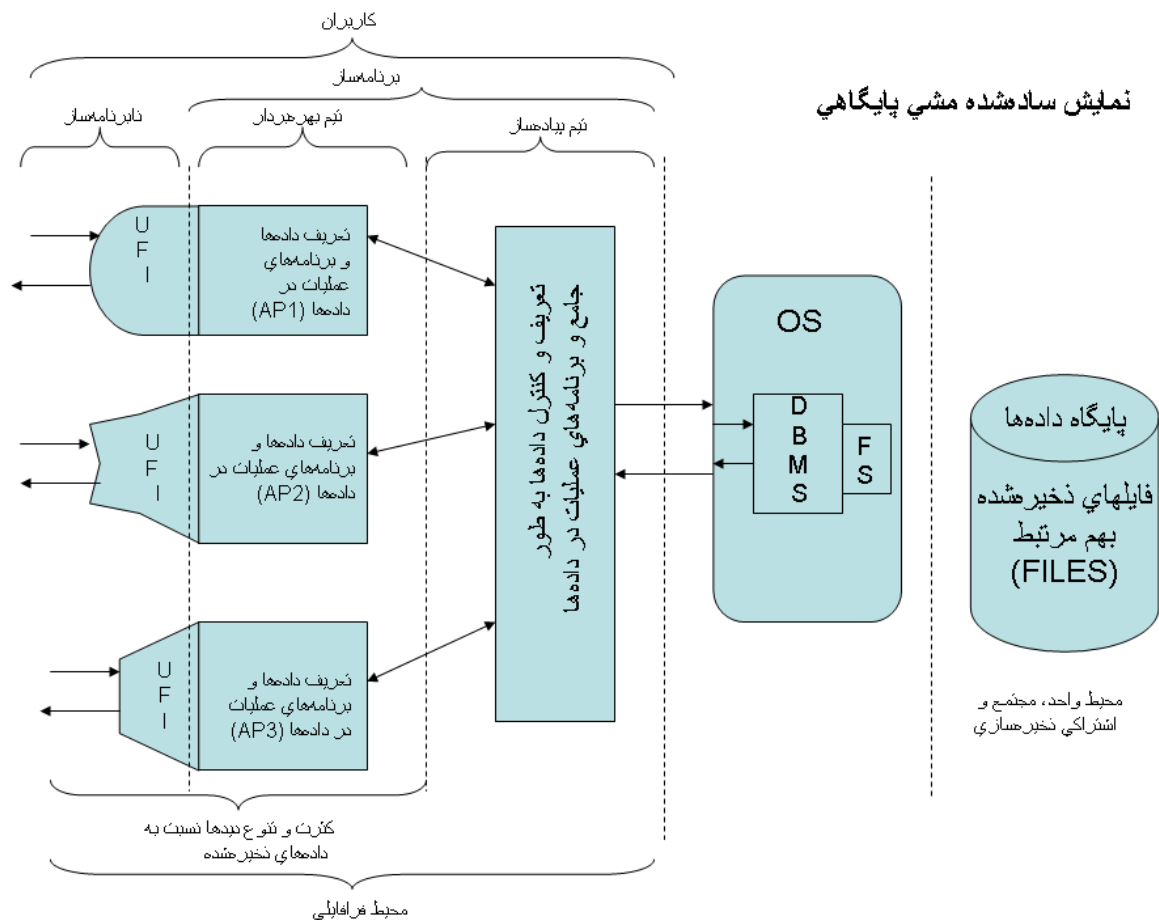
- اجرای مراحل کلاسیک اولیه لازم برای طراحی و تولید یک سیستم کاربردی

- 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.

۸- معایب مشی فایلینگ

- ۱- عدم وجود محیط مجتمع ذخیره‌سازی اطلاعات و عدم وجود سیستم یکپارچه
- ۲- عدم وجود سیستم کنترل متمرکز روی کل داده‌ها
- ۳- افزونگی
- ۴- عدم وجود ضوابط ایمنی کارا و مطمئن
- ۵- خطر بروز پدیده ناسازگاری داده‌ها
- ۶- عدم امکان اشتراکی شدن داده‌ها
- ۷- مصرف نابیهینه امکانات سخت‌افزاری و نرم‌افزاری
- ۸- حجم زیاد برنامه‌سازی
- ۹- وابستگی برنامه‌های کاربردی به محیط ذخیره‌سازی داده‌ها

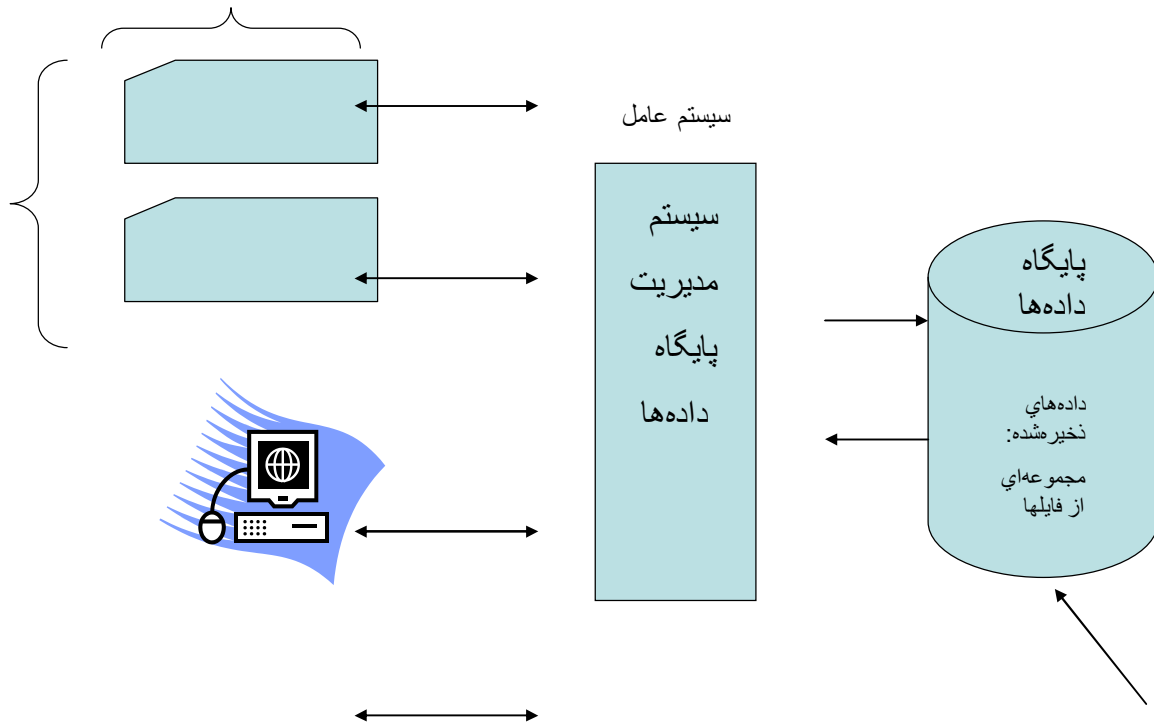


۹- مراحل کلی کار در مشی پایگاهی

- بررسی و تحلیل نیازهای پردازشی و اطلاعاتی همه قسمتها توسط يك گروه
- تعیین مشخصات جامع (یکپارچه) کاربردی و وظایف آن
- انتخاب يك یا چند پیکربندی سخت افزاری-نرم افزاری
- استفاده از يك یا چند DBMS
- طراحی پایگاه دادهها در سطوح لازم
- تولید مجموعه ای از برنامه های ایجاد و کنترل پایگاه داده
- ایجاد محیط واحد و مجتمع ذخیره سازی و مشترك بين کاربران
- طراحی و تولید واسطه های کاربر پسند مورد نیاز
- تعریف پایگاه داده هر قسمت توسط کاربر مربوطه
- طراحی برنامه های عملیاتی در پایگاه داده
- بهره برداری واقعی از سیستم پس از تستهای لازم

۱۰- عناصر محیط پایگاه داده‌ها

۱- سخت‌افزار ۲- نرم‌افزار ۳- کاربر ۴- داده



۱۱- انواع سخت‌افزارهای محیط پایگاه داده

سخت‌افزار ذخیره‌سازی داده‌ها

سخت‌افزار پردازشگر

سخت‌افزار ارتباطی

۱۲- انواع نرم‌افزارهای موجود در محیط پایگاه داده‌ها

سیستم مدیریت پایگاه داده‌ها (DBMS)

برنامه‌های کاربردی قابل اجرا در محیط DBMS

رویه‌های ذخیره‌شده

نرم‌افزار شبکه

مدلسازی معنایی داده‌ها

۱- مدلسازی معنایی داده‌ها

داده‌های ذخیره‌شده در پایگاه داده‌ها ابتدا باید در بالاترین سطح انتزاع مدلسازی معنایی شوند.

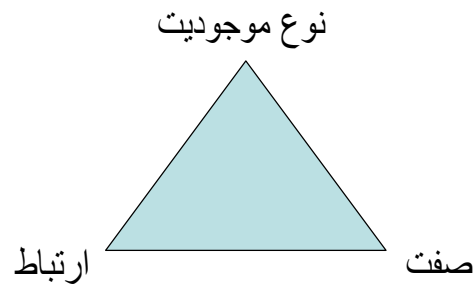
۲-انواع روشهای مدلسازی معنایی داده‌ها

روش موجودیت-ارتباط (ER)

روش زبان عمومی مدلسازی (UML)

روش تکنیک مدلسازی شیئی (OMT)

۳-سه مفهوم معنایی موجود در روش ER



۴-تعریف موجودیت

مفهوم کلی شیئی، چیز، پدیده و به طور کلی هر آنچه که می‌خواهیم در موردش اطلاع داشته باشیم و شناخت خود را در موردش افزایش دهیم.

۵-سه ضابطه در رابطه با تشخیص یک نوع موجودیت

۱- معمولاً نمونه‌هایی متمایز از یکدیگر دارند.

۲- معمولاً بیش از یک صفت دارد و کاربر به مجموعه‌ای از اطلاعات در مورد آن نیاز دارد.

۳- معمولاً حالت کنشگری (فاعلیت) یا حالت کنشپذیری (مفعولیت) دارد.

۶-موجودیت مستقل و وابسته

موجودیت مستقل (قوی)، موجودیتی است که مستقل از هر موجودیت دیگر و به خودی خود، در یک محیط مشخص مطرح باشد.

موجودیت وابسته (ضعیف)، موجودیتی است که وجودش وابسته به یک نوع موجودیت دیگر است.

۷-تعریف صفت

خاصیت یا ویژگی یک نوع موجودیت است و هر نوع موجودیت مجموعه‌ای از صفات دارد. هر صفت یک نام، یک نوع و یک معنای مشخص دارد.

۸-انواع صفت

ساده یا مرکب

تکمقداری یا چندمقداری

شناسه یا ناشناسه موجودیت

هیچ مقدارپذیر یا ناپذیر

ذخیره شده یا مشتق

مقدار صفت ساده از لحاظ معنایی تجزیه نشدنی یا اتمیک است.

صفت مرکب از چند صفت ساده تشکیل شده است.

صفت تک‌مقداری، صفتی است که برای یک نمونه از یک نوع موجودیت حداکثر یک مقدار از دامنه مقادیر را می‌گیرد

صفت چندمقداری بیش از یک مقدار از دامنه مقادیر می‌گیرد.

صفت شناسه موجودیت که گاه به آن کلید هم گفته می‌شود، دو ویژگی دارد:

۱- یکتایی مقدار دارد.

۲- حتی الامکان طول مقادیرش کوتاه است.

صفت هیچ مقدارپذیر

هیچ مقدار یعنی مقدار ناشناخته، مقدار غیرقابل اعمال، مقدار تعریف نشده.

اگر مقدار یک صفت در یک یا بیش از یک نمونه از یک نوع موجودیت، برابر با هیچ مقدار باشد، آن صفت هیچ مقدارپذیر است.

صفت ذخیره شده و مشتق

صفت ذخیره شده صفتی است که مقادیرش در پایگاه داده‌ها ذخیره شده باشد.

صفت مشتق، صفتی است که مقادیرش در پایگاه داده‌ها ذخیره نشده باشد، بلکه حاصل یک پردازش روی فقره‌هایی از داده‌های ذخیره شده باشد. تعامل بین دو یا بیش از دو نوع موجودیت است و ماهیت آن نوعی بستگی بین نوع-موجودیتها است.

خصوصیات نوع ارتباط

۱- هر ارتباط یک نام دارد

۲- هر ارتباط یک معنای مشخص دارد و این معنا با معنای هر ارتباط دیگر متفاوت است.

۳- هر ارتباط نمونه‌هایی دارد.

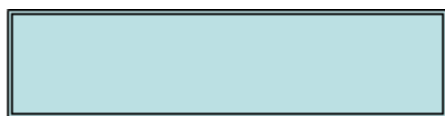
نمودار ER

نموداری است که سه مفهوم اساسی مدل ER، یعنی نوع موجودیت، صفت و ارتباط نمایش داده می‌شوند

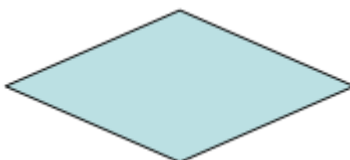
نوع موجودیت



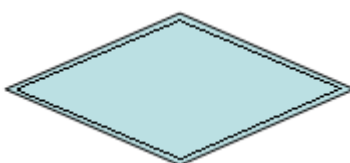
نوع موجودیت ضعیف (وابسته)



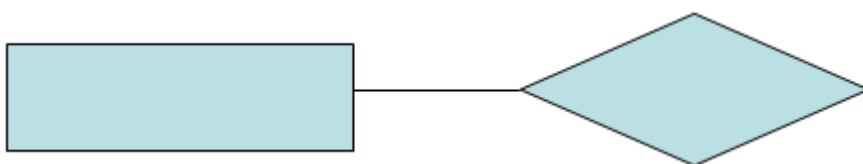
نوع ارتباط



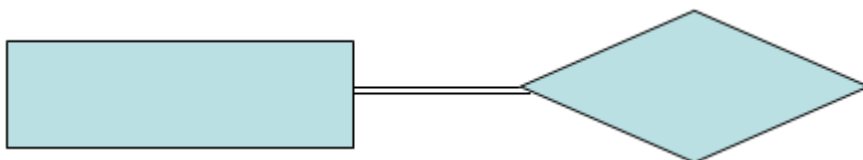
نوع ارتباط موجودیت ضعیف با قوی



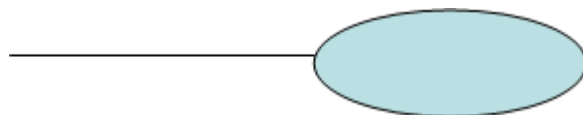
مشارکت نوع موجودیت در نوع ارتباط



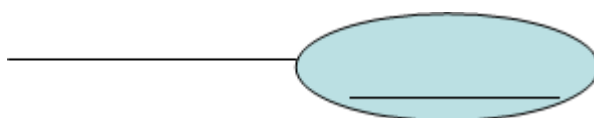
مشارکت الزامی



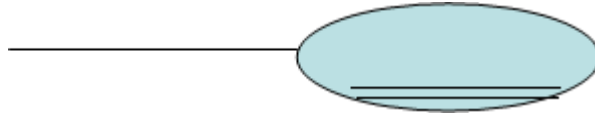
صفت



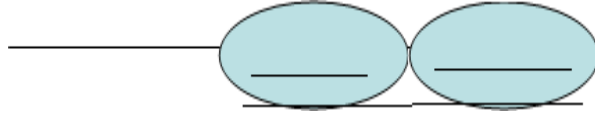
صفت شناسه اول



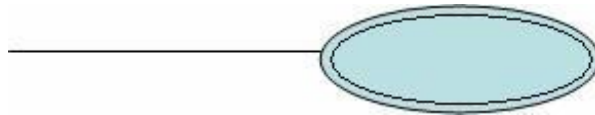
صفت شناسه دوم (در صورت وجود)



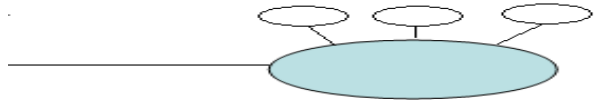
صفت شناسه مرکب (مثلا دو صفتی)



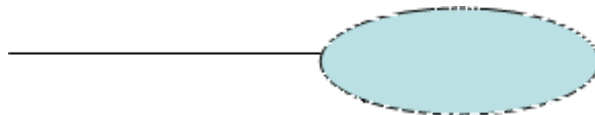
صفت چند مقداری



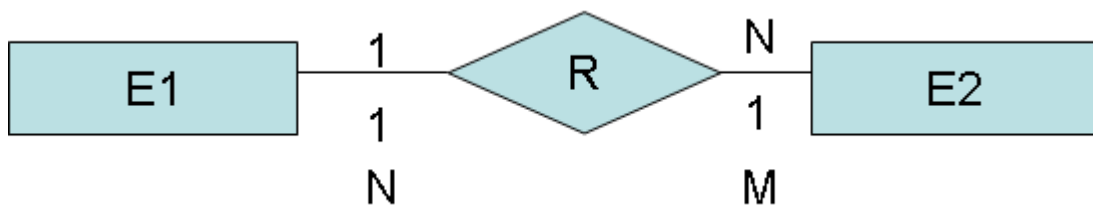
صفت مرکب



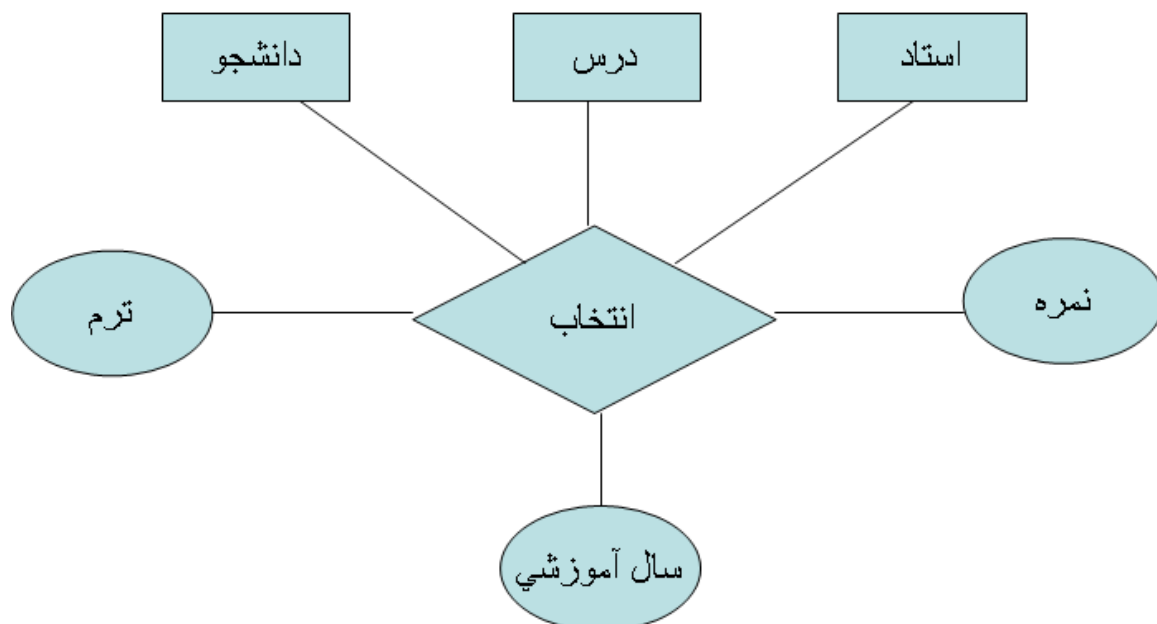
صفت مشتق (محاسبه شدنی)



چند ارتباط



تعداد شرکت کنندگان در يك ارتباط را درجه آن ارتباط مي گویند.



ارتباط بین سه موجودیت

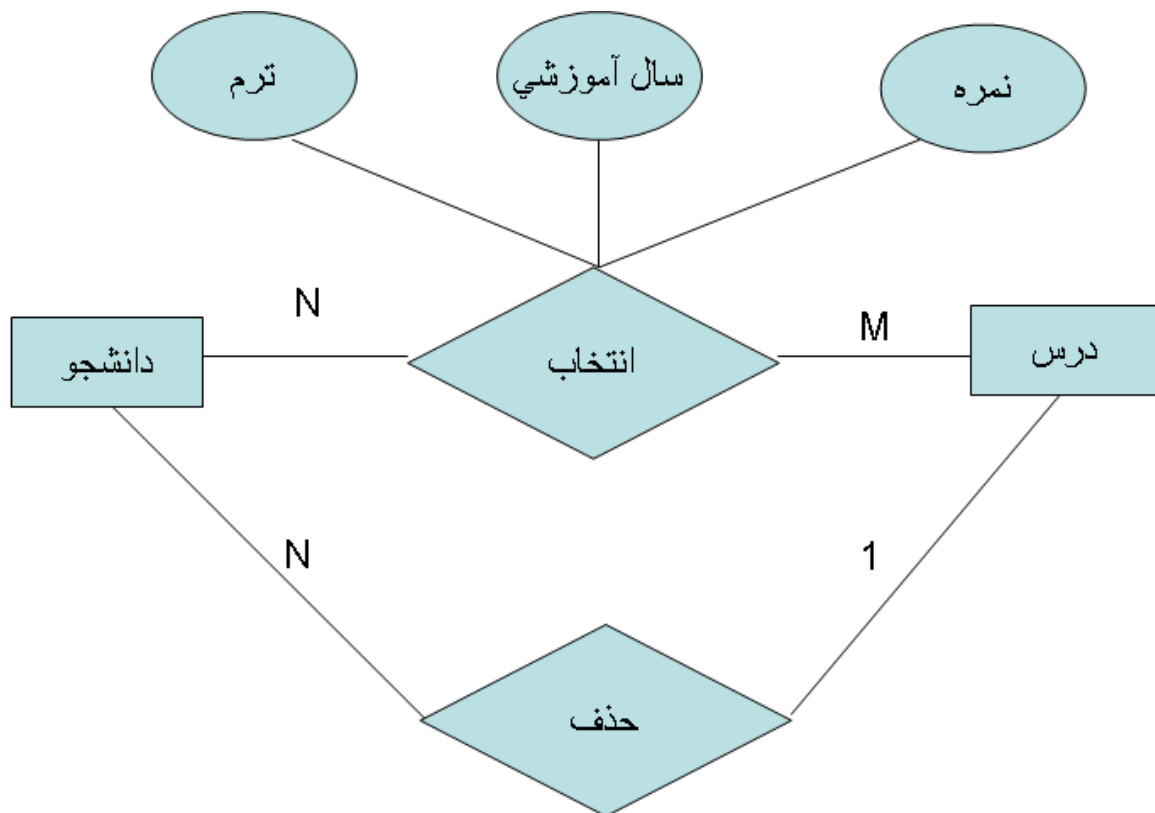
چندی یا ماهیت نوع ارتباط عبارتست از چگونگی تناظر بین دو مجموعه نمونه‌های آن دو نوع موجودیت.

یک به یک 1:1

یک به چند 1:N

چند به چند N:M

نمایش چندی ارتباط



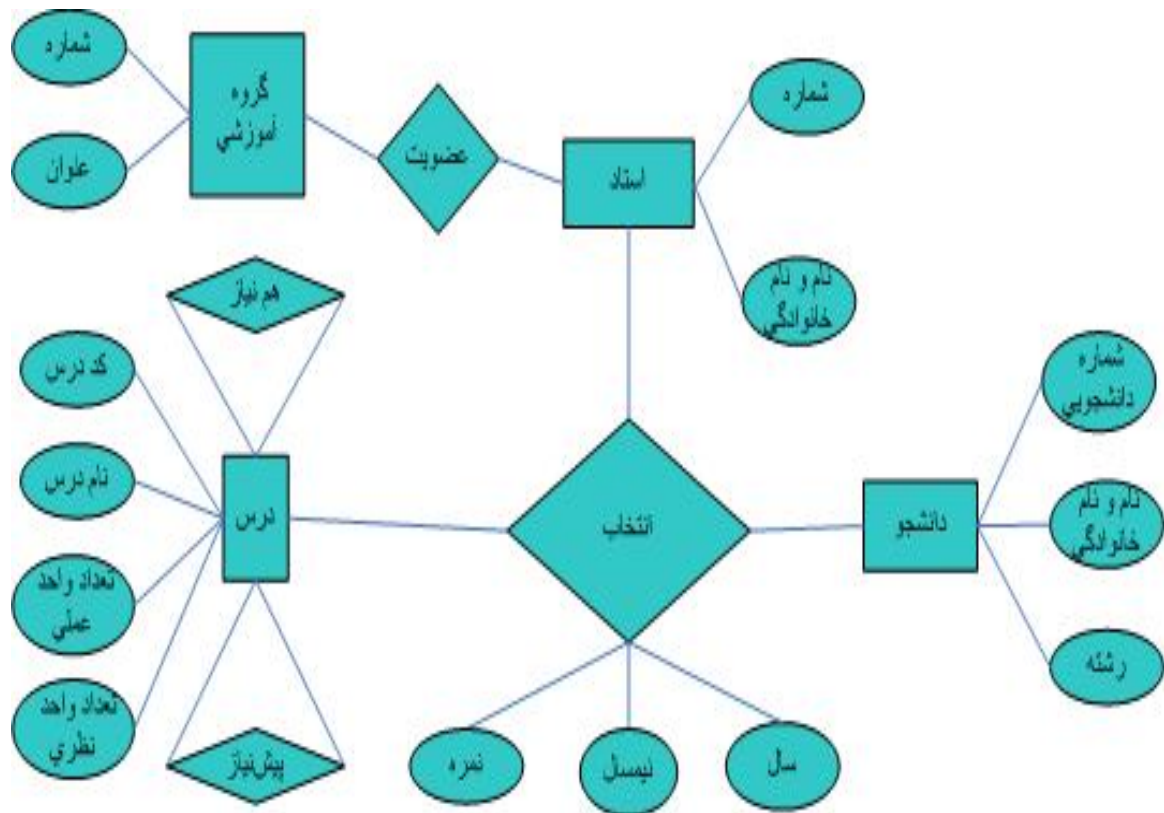
مثال : مدلسازي بسيار ساده شده داده‌هاي دانشكده

بخشي از يك نمودار ER ممكن براي قسمتي از داده‌هاي اين محيط را رسم مي‌كنيم. موجوديتهاي مطرح عبارتند از :

*دانشجو *درس *اسناد *گروه آموزشي

تمرين :

۱. مجموعه صفات هر نوع موجوديت و وضع هر صفت را مشخص كنيد (ساده، مركب، تك مقداري، چند مقداري ...)
۲. الزامي بودن يا نبودن مشاركت در هر نوع ارتباط را در نمودار نشان دهيد (در صورت لزوم)
۳. هر جا لازم شد، صفات انواع ارتباط را تعيين كنيد.
۴. چندي هر ارتباط را مشخص كنيد.
۵. انواع ديگري از موجوديتها و ارتباطات را به آن اضافه كنيد.
۶. مجموعه پرسش (پرس و جو) هايي كه اين نمودار مي‌تواند به آنها پاسخ دهد را فهرست كنيد.



مدل رابطه‌ای

بخش‌های اساسی مدل داده‌ای

۱- بخش ساختاری

بخش ساختاری، نشان‌دهنده عناصر ساختاری مدل است که همان ساختار داده‌ای اصلی و مفاهیم مرتبط با آن است.

۲- بخش عملیاتی (پردازشی)

بخش عملیاتی، مجموعه امکاناتی است که به وسیله آنها عملیات مورد نظر کاربر انجام می‌شود.

۳- بخش جامعیتی

بخش جامعیتی، از مجموعه‌ای از قواعد و محدودیت‌های جامعیتی تشکیل شده است که به وسیله آنها سیستم مدیریت پایگاه داده می‌تواند صحت، دقت و سازگاری داده‌ها را کنترل و تضمین کند.

تعریف رابطه

با فرض وجود n میدان D_1 تا D_n ، نه لزوماً متمایز، رابطه R از دو قسمت تشکیل شده است:

۱- سرآیند: مجموعه‌ای نامدار از n صفت به صورت $A_i: D_i$ که در آن هر A_i نام یک صفت است و هر D_i نام میدان صفت

۲- پیکر (بدنه): مجموعه‌ای است از m تاپل t به نحوی که t خود مجموعه‌ای است از n عنصر هر یک به صورت $A_i: v_i$ که در آن v_i مقداری است از نوع میدان.

مقدار n را درجه (همان تعداد صفات) و مقدار m را کاردینالیته رابطه می‌گویند

تناظر بين مفاهيم رابطه اي و مفاهيم جدولی

جدول زیر تناظر بين مفاهيم رابطه اي و مفاهيم جدولی را نشان مي دهد :

مفهوم جدولی	مفهوم تئوريك
جدول	رابطه
سطر	تاپل
ستون	صفت
مجموعه مقادير ستون	ميدان
تعداد ستونها	درجه
تعداد سطرها	كاردینالیتی

ویژگیهای رابطه

۱- رابطه تاپل تكراري ندارد

۲- تاپلها نظم ندارند

۳- صفات رابطه نظم مكاني ندارند

۴- تمام صفات تكمقداري هستند.

انواع رابطه

۱- مبنا: استقلال وجودی دارد و از رابطه‌های دیگر مشتق نیست و داده‌های ذخیره‌شده متناظر دارد.

۲- نامدار: با يك نام به سیستم معرفی می‌شود.

۳- مشتق: به كمك يك عبارت رابطه‌اي بر حسب رابطه‌های مبنا تعریف می‌شود.

۴- عبارتی: از مجموعه‌اي از رابطه‌های نامدار و به وسیله يك عبارت رابطه‌اي به دست می‌آید.

۵- نتیجه پرسش: بی‌نام و مشتق است که حاصل اجرای يك پرسش مشخص است.

۶- مشتق: رابطه‌اي عبارتی است که به طور مستقیم و کارا ذخیره شده است.

میدان (دامنه)

از نظر ریاضی، مجموعه‌اي است از مقادير که يك یا بیش از يك صفت از آن مقدار می‌گیرند. در مدل رابطه‌اي، این مجموعه:

-نامدار است.

-مقاديرش نوع مشخص دارند.

-مقاديرش فرمت مشخص دارند.

مزایای میدان

۱. سبب ساده‌تر شدن و کوتاه‌تر شدن شمای پایگاه داده‌ها می‌شود.

۲. امکانی است برای کنترل مقداری عملیات در پایگاه داده‌ها.

۳. امکانی است برای کنترل معنایی درخواستها.

۴. پاسخگویی به بعضی پرسشها را آسان می‌کند.

۵. با استفاده از مفهوم میدان می‌توان امکانات و جنبه‌های موجود در سیستم مدیریت پایگاه داده‌های شیئی‌گرا را به سیستم مدیریت پایگاه داده‌های رابطه‌ای افزود.

رابطه نرمال و غیر نرمال

رابطه نرمال :

رابطه‌ای است که مقادیر هیچیک از میدانهایش ، یک مقدار رابطه‌ای با کاردینالیتی بزرگتر از یک نباشد

رابطه غیر نرمال :

رابطه‌ای است که در آن مقادیر حداقل یک صفت ، خود مقادیر رابطه‌ای هستند

(رابطه ۲_ای که حداقل یک صفت چند مقداری داشته باشد)

دلیل لزوم نرمال بودن رابطه

۱- سادگی در نمایش ظاهری رابطه (جدول با سطرهای ساده‌تر)

۲- سادگی دستورات DSL (بویژه DML و DDL)

۳- سادگی در اجرای عملیات در پایگاه داده‌ها

معایب رابطه نرمال

۱- بروز پدیده افزونگی که می‌تواند فیزیکی هم باشد.

۲- طولانی‌تر شدن کلید رابطه.

۳- عدم امکان نمایش داده‌های پیچیده.

۴- دشواری در نمایش طبیعی مفهوم سلسله مراتب.

۵- دشواری در نمایش مفهوم وراثت.

مزایا و معایب رابطه غیرنرمال

مزایا:

۱- کاهش میزان افزونگی

۲- کوتاه شدن کلید

۳- امکان نمایش داده‌های پیچیده

۴- دشواری کمتر در نمایش مفهوم سلسله‌مراتب و مفهوم وراثت

۵- افزایش سرعت عملیاتی سیستم در بازیابی اطلاعات

۶- عدم نیاز به نرمالترسازی رابطه

معایب:

۱- پیچیدگی

۲- عدم تقارن صفات

کلید در مدل رابطه‌ای

۱. super key / ابر کلید
۲. candidate key / کلید کاندید
۳. primary key / کلید اصلی
۴. alternate key / کلید بدیل
۵. foreign key / کلید خارجی

ابر کلید

هر زیرمجموعه از مجموعه عنوان رابطه که یکتایی مقدار در گستره رابطه داشته باشد.

کلید کاندید

هر زیرمجموعه از مجموعه عنوان رابطه که دو خاصیت زیر را داشته باشد کلید کاندید رابطه است:

۱- یکتایی مقدار

۲- کاهش ناپذیری

کلید اصلی

یکی از کلیدهای کاندید رابطه که طراح انتخاب می‌کند و به سیستم معرفی می‌شود.

ضابطه‌های انتخاب:

۱- از نظر کاربرد، شناسه معمول نوع موجودیت باشد.

۲- طول کوتاهتر داشته باشد.

کلید دیگر (بدیل)

هر کلید کاندید، غیر از کلید اصلی، کلید دیگر نام دارد.

کلید خارجی

دورابطه R1 و R2 را در نظر می‌گیریم. هر زیرمجموعه از صفات رابطه R2 که هر مقدار معلومش با یک مقدار از کلید کاندید R1 برابر باشد، کلید خارجی در رابطه R2 است.

کلید خارجی برای نمایش ارتباطات بین انواع موجودیتها به کار می‌رود.