

بسمه تعالی

مجموعه آموزش‌های مجازی دانشگاه معین رشت

طراحی بانک‌های اطلاعاتی

جلسه اول

مدرس سید علی اصغری

منابع:

- Database System Concepts, Abraham Silberchatz, Henry F.Korth
- اصول طراحی پایگاه داده‌ها، آبراهام سیلبرشاتز، ترجمه جعفرنژاد قمی
- سیستم‌های بانک اطلاعاتی، سی جی دیت، ترجمه امی علیخانزاده

منبع اصلی مطالبی که ارائه می‌شود کتاب سیلبرشاتز بوده که توسط جعفرنژاد ترجمه شده است. با توجه به تعطیلی دانشگاه‌ها و عدم تشکیل کلاس حضوری توصیه می‌شود که حتماً مطالب درسی از کتاب مطالعه شوند. در صورتی که منابع آنلاین مدنظر بود در همین سری آموزش‌ها در لایه لای دروس به دانشجویان معرفی خواهد شد.

برای ارتباط با مدرس می‌توانید از طرق زیر به ترتیب اولویت اقدام نمایید. لطفاً در هر ایمیلی که ارسال می‌کنید خودتان را بطور کامل معرفی کنید.

saasghari@chmail.ir

saasghari@gmail.com

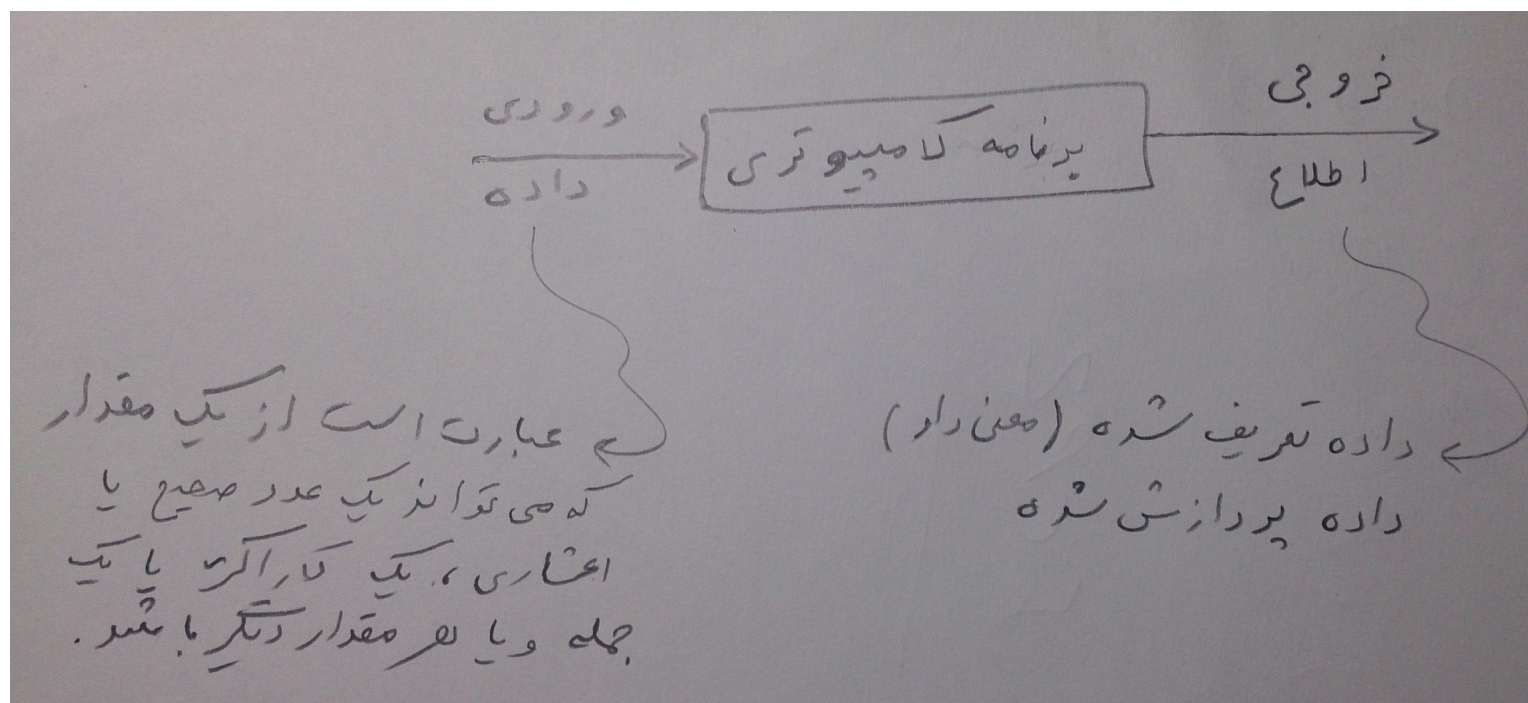
(فقط واتس اپ) ۰۹۱۱۲۳۲۹۰۱۱

مطالب درس:

- مقدمه‌ای بر پایگاه داده‌ها
- معرفی پایگاه داده رابطه‌ای
- انواع کلیدها
- جبر رابطه‌ای و عملگرهای مختلف آن
- زبان SQL و نحوه استفاده از آن در پایگاه داده
- طراحی پایگاه داده و مدل ER
- وابستگی‌های تابعی
- فرم‌های نرمال پایگاه داده

داده و اطلاع:

هر نرم افزار در کامپیوتر با داده ها و اطلاعات سرو کار دارد و می توان هر برنامه را به شکل زیر در نظر گرفت



داده عبارت است از یک مقدار. این مقدار میتواند یک عدد صحیح یا اعشاری، یک کاراکتر (حرف یا نماد)، یک رشته یا دنباله ای از کاراکترها، به صورت یک کلمه یا جمله و یا هر چیز دیگری که یک مقدار محسوب میشود باشد.

داده میتواند شامل مقادیر ساده باشد که به آن داده جزء یا عنصر پایه میگویند. مثل 15 یا «بهمن» که غیر قابل تجزیه میباشد. در مقابل عنصر پایه داده‌هایی هستند که از چند داده جزء تشکیل شده‌اند که به آنها عناصر گروهی یا داده‌های گروهی گفته میشود. هر عنصر گروهی از چند زیر عنصر تشکیل میشود که هر کدام میتوانند یک داده جزء یا داده گروهی باشد. مانند «15 بهمین 1384» که یک داده گروهی بوده و تاریخ یک روز خاص را مشخص مینماید. این داده گروهی شامل سه قسمت است. یکی برای روز یکی برای ماه و دیگری برای مشخص کردن سال میباشد. هر کدام از این سه داده یک زیر عنصر از عنصر گروهی تاریخ میباشد.

اطلاع عبارت است از داده‌ی پردازش شده یا داده‌ی معنی دار.

مثلاً 15 تنها یک عدد است و نمیتوانیم چیز دیگری در باره آن بگوییم. ولی وقتی مینویسیم: «سن 15»، یا مینویسیم: «نمره 15»، کاملاً مشخص کرده‌ایم که مقصودمان از 15 چیست. 15 به تنهایی یک داده است و وقتی با کلمه «نمره» یا کلمه «سن» همراه میشود معنی و مفهومی پیدا میکند و به آن اطلاع میگوییم. ما اطلاع داریم که نمره 15 است و یا سن 15 است. یا مثلاً کلمه «نرگس» را در نظر بگیرید. این کلمه تنها یک مقدار یا یک داده است و مشخص نیست که مقصود نام یک شخص است یا نام یک گل. اما وقتی مینویسیم «گل نرگس» کاملاً منظورمان را مشخص میکنیم.

حافظه کامپیوتر تنها می‌تواند اعداد دودویی را در خود ذخیره کند. در همین حافظه است که همه انواع داده‌ها را با روش‌های کد گذاری به صورت عددی ذخیره می‌کنیم. بنابراین اگر در جایی از حافظه عدد ۶۴ ذخیره شده باشد این عدد ممکن است حرف A باشد یا کد دستورالعملی که در پردازنده باید اجرا شود، یا یک شماره رنگ یکی از نقاط روی یک عکس و یا فرکانس صوتی که باید از اسپیکر پخش شود. یا فقط یک عدد صحیح ۶۴ که قرار است در یک برنامه در یک عبارت محاسباتی مورد استفاده قرار گیرد. توجه داشته باشید که همه جا مقدار داده یکسان و برابر است با عدد ۶۴. اما در جاهای مختلف همین عدد مفاهیم متفاوتی دارد و پردازنده بر اساس ساختمان داده تعریف شده و محل ذخیره سازی داده اطلاع می‌یابد که چگونه باید از آن استفاده نماید.

مفاهیم داده و اطلاع از مفاهیم اساسی در کامپیوتر هستند که برخلاف تعریف ظاهراً ساده و پیش پا افتاده‌ای که دارند در بسیاری از مباحث از جمله در پایگاه داده درک مفاهیم زیادی در گرو درک مفهوم داده و اطلاع و تمایز آنها است.

ضرورت بحث پایگاه داده

هر نرم افزار در کامپیوتر با داده ها و اطلاعات سر و کار دارد. وقتی داده ها برای پردازش به برنامه کامپیوتری وارد می شوند این برنامه ها ممکن است لازم باشد که داده ها را در جایی روی دیسک یا هر نوع حافظه جانبی دیگر ذخیره کند. می توان برای این کار به سادگی از یک فایل برای ذخیره داده ها استفاده کرد. این کار برای یک برنامه کوچک و زمانی که حجم داده ها کم باشد ممکن است با اشکالی مواجه نشود. اما اگر حجم داده ها زیاد باشد و یا قرار باشد یک نرم افزار بزرگ به صورت تیمی ساخته شود با اشکالاتی بدین شرح مواجه می شویم:

- **ناسازگاری و افزونگی داده ها:** بدلیل تعدد برنامه نویسان و طولانی بودن پروسه توسعه نرم افزار ممکن است فایل های داده ساختارهای متفاوت داشته باشند. ممکن است اطلاعات یکسان در چندین محل ذخیره (تکرار) شوند که به آن افزونگی داده می گوئیم. از طرفی این کپی های متعدد از داده های یکسان ممکن است در زمان به روز رسانی به هر دلیلی با هم سازگار نباشند. یعنی بعضی به روز شده بعضی به روز نشوند. به این حالت ناسازگاری داده می گوئیم.

- **دشواری دستیابی به داده‌ها:** در صورتی که نیازهای نرم‌افزاری جدیدی مورد تقاضا باشد و لازم باشد گزارش‌گیری‌هایی توسط نرم‌افزار ارائه شود که قبلاً نبوده یا باید به صورت دستی انجام شود و یا وسط برنامه‌نویسان برنامه جدیدی نوشته شود و به سیستم اضافه گردد که بسیار وقت‌گیر و هزینه‌بردار است.
- **تفکیک داده‌ها:** بدلیل تعدد فایل‌ها و تفاوت ساختار آنها نوشتن برنامه‌های کاربردی جدید برای دستیابی به داده‌های مناسب دشوار است.
- **مشکلات جامعیت:** برای کنترل داده‌ها لازم است که قیدهای سازگاری روی داده‌ها قرار داده شود. مثلاً کنترل شود که همیشه نمره مقداری بین صفر تا بیست داشته باشد. در سیستم‌های پردازش فایل این کار با کدنویسی و در داخل برنامه انجام می‌شود. ممکن است لازم باشد در جاهای مختلف برنامه شرط‌هایی قرار داده شود. اعمال قیود جدید می‌تواند بسیار پیچیده بوده و احتمال اشتباه در آنها زیاد است.

- **مشکلات یکپارچگی (اتمیک):** هر عمل در پایگاه داده و هر تغییر در داده‌های آن مجموعه‌ای از ریز عملیات میباشد. در هنگام انجام يك عمل در پایگاه داده ممکن است تعدادی از این ریز عملیات انجام شده و کار تا نیمه انجام شده باشد و به هر دلیلی عملیات متوقف شود. مثلاً برق قطع شود. در چنین شرایطی مشخص نیست که تا کجای کار انجام شده و ادامه کار از کجا باید از سر گرفته شود. این مسأله باعث خرابی داده‌ها میشود. برای رفع مشکل باید هر عملیات پایگاه داده (تراکنش) به صورت یکپارچه یا اتمیک انجام شود. یعنی یا کل عملیات کاملاً انجام میشود یا اینکه اصلاً شروع نمیشود.

- **ناهنجاری دستیابی همزمان:** در پایگاه داده ممکن است چند کاربر به طور همزمان به پایگاه دسترسی داشته باشند و هر دو با هم به يك محل از پایگاه مراجعه نمایند. این مراجعه همزمان ممکن است منجر به ناسازگاری داده‌ها گردد.

- **مشکلات حفاظتی:** هر کسی نمیتواند به هر جایی از پایگاه داده دسترسی داشته باشد و دسترسیها باید محدود گردد.

حتی زمانی که برنامه بزرگ نباشد و حجم داده‌ها زیاد نباشد نیز مشکلاتی داریم. مطمئناً برای ذخیره و بازیابی اطلاعات از حافظه لازم است که ساختمان داده مناسبی برای فایل انتخاب یا طراحی گردد. از طرفی الگوریتم‌های مربوطه به ذخیره و بازیابی داده‌ها، درج در ابتدا یا وسط، جستجو و به روز رسانی باید پیاده‌سازی گردد. از طرفی برای جداسازی داده‌ها در داخل فایل و مشخص کردن پایان هر فیلد و هر رکور باید از علایم و قرار دادهای خاصی تبعیت شود.

اگر لازم باشد برای هر برنامه کوچک این موارد و بسیاری از موارد دیگر که در اینجا ذکر نشده طراحی و پیاده‌سازی شود هزینه ساخت نرم‌افزارها بسیار زیاد خواهد بود. اینجاست که می‌گوییم «چرخ قبلاً اختراع شده». تمام آنچه که نرم‌افزار برای ذخیره و بازیابی سریع نیاز دارد در نرم‌افزارهای با عنوان DBMS پیاده‌سازی شده‌اند.

پایگاه داده چیست؟

به محل نگهداری داده‌ها یا کلکسیون داده‌ها پایگاه (DB) داده گفته میشود. پایگاه داده را میتوان به يك بایگانی تشبیه کرد که در آن اطلاعات و داده‌ها نگهداری میشود.

سیستم مدیریت پایگاه داده (DBMS) مجموعه داده‌های مرتبط به هم و برنامه‌های کامپیوتری است که برای دستیابی و دستکاری آن داده‌ها میباشد. سیستم بانک اطلاعاتی به طور عمومی يك سیستم کامپیوتری نگهداری رکوردها میباشد.

سیستم مدیریت، پایگاه داده داده‌ها را نگهداری کرده و امکان دستیابی و دستکاری آنها را بر اساس ضوابط مشخص شده برای کاربر فراهم مینماید.

هدف از طراحی و پیاده‌سازی سیستم‌های بانک اطلاعاتی مدیریت حجم زیادی از داده‌ها میباشد. این مدیریت شامل تعریف و ارائه‌ی ساختارهایی برای ذخیره داده‌ها و همچنین راهکارهایی برای دستکاری و امنیت (مسأله‌ی محافظت اطلاعات و محدودیت‌های دسترسی) میباشد.

