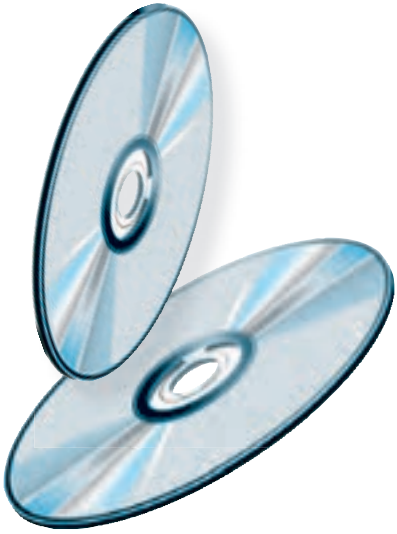
هدف کلی

آشنایی با مفاهیم بانک اطلاعاتی و کار با DBMS متداول



فصل ۱

مفاهیم بانک اطلاعاتی[۱]

در این فصل به معرفی تعاریف و مفاهیم اولیهٔ بانک‌های اطلاعاتی خواهیم پرداخت. اگرچه نرم‌افزارهای کاربردی به‌سهولت قابل استفاده هستند، اما بدون دانستن دقیق مفاهیم نمی‌توان از آن‌ها به‌طور مناسب استفاده کرد و هیچ‌گاه نمی‌توان بانک اطلاعاتی کارآمدی را ایجاد و پیاده‌سازی کرد.

هدف‌های رفتاری: پس از آموزش این فصل هنرجو می‌تواند:

- **انواع سیستم‌های ذخیره و بازیابی را توضیح دهد.**
- **مفهوم فیلد، رکورد، فایل را توضیح دهد.**
- **بانک اطلاعاتی را تعریف کند.**
- **سیستم مدیریت بانک اطلاعاتی و معماری آن را شرح دهد.**
- **مفاهیم مدل رابطه‌ای را شرح دهد.**
- **سیستم بانک اطلاعاتی را طراحی کند.**

۱ـ۱ـ کلیات

بانک‌های اطلاعاتی یکی از انواع سیستم‌های «ذخیره و بازیابی اطلاعات»[۲] است. سیستم ذخیره و بازیابی سیستمی است که به کاربر امکان می‌دهد تا داده‌ها و اطلاعات خود را ذخیره، بازیابی و پردازش نماید.

۱ــ در این کتاب پایگاه داده‌ها، بانک اطلاعاتی، بانک داده و بانک به یک معنی استفاده شده است.

۲ــ Information Storage and Retrieval

نظر به تعریف عام بالا، موارد زیر قابل ذکر است :

۱ــ کاربر می‌تواند دارای قابلیت «برنامه‌سازی» یا فاقد این مهارت باشد.

۲ــ داده می‌تواند متن، تصویر، صوت و نظایر آن باشد.

سیستم ذخیره و بازیابی اطلاعات با گذشت زمان، از مفهوم سادهٔ فایلینگ[۱] به‌صورت زیر توسعه پیدا کرده است :

۱ــ سیستم فایلینگ Filing System (FS)

۲ــ سیستم مدیریت داده‌ها Data Management System (DMS)

۳ــ سیستم مدیریت پایگاه داده‌ها Database Management System (DBMS)

۴ــ سیستم مدیریت پایگاه دانش Knowledge Base Management System (KBMS)

پیش از تعریف سیستم مدیریت پایگاه داده بهتر است، برخی از مفاهیم، تعاریف مقدماتی و مفهوم سیستم فایلینگ بررسی شود تا با دقت بیشتری بتوانیم مفهوم بانک اطلاعاتی را بشناسیم.

۲ـ۱ـ مفاهیم مقدماتی در ذخیره و بازیابی

۱ــ۲ــ۱ــ فیلد[۲] : یک قطعه داده است[۳].

هر فیلد دارای دو جزء می‌باشد :

۱ــ نام

۲ــ مقدار

مقدار، حکم داده[۴] را دارد و نام فیلد به همراه مقدار در حکم اطلاع[۵] خواهد بود. به‌عنوان **مثال :**

نام‌خانوادگی = علوی

نام فیلد — مقدار فیلد

در واقع هر فیلد یک ویژگی یا صفت[۶] به‌حساب می‌آید. توجه کنید که از نام فیلد و مقدار فیلد به همراه هم اطلاع پیدا می‌کنیم که نام‌خانوادگی فرد موردنظر علوی است. مثال جالب‌تر این‌که : مقدار عددی «۸۸۹۶۵۴۲۳» هیچ معنی خاصی ندارد. حال به این عبارت توجه کنید : «تلفن = ۸۸۹۶۵۴۲۳»،

۱ــ Filing ۲ــ Field

۳ــ فیلد واحد معنایی داده است. (Semantic Data Unit)

۴ــ Data ۵ــ Information ۶ــ Attribute

با درج نام فیلد در کنار مقدار عددی، شما اطلاع پیدا می‌کنید که آن عدد شماره تلفن است. اگر به‌جای تلفن، شماره شناسنامه قرار گیرد اطلاع کسب شده متفاوت خواهد بود.

نکته

یکی از خصوصیات مهم فیلد، نوع داده‌ای است که در آن قرار می‌گیرد. مثلاً Integer ، Real یا Text و... .

۲ــ۲ــ۱ــ رکورد[1] : به مجموعهٔ فیلدهای مرتبط، یک رکورد می‌گوییم.

به‌عنوان مثال، قالب رکورد مربوط به یک دانش‌آموز می‌تواند شامل نام فیلدهای زیر باشد :

نام، نام‌خانوادگی، سال تولد، مقطع تحصیلی، کلاس و...

محتوای رکوردهای دانش‌آموزان می‌تواند برای دو دانش‌آموز همکلاسی به‌صورت زیر باشد :

۱ــ مجید، شجاعی، ۱۳۷۰، اول دبیرستان، کلاس ۱ الف و...

۲ــ حسن، کمالیان، ۱۳۷۱، اول دبیرستان، کلاس ۱ الف و...

محتوای رکورد می‌تواند دارای طول ثابت یا متغیر باشد. درحالت رکورد با طول ثابت، تعداد، ترتیب و اندازهٔ فیلدهای متناظر در تمام رکوردها یکسان درنظر گرفته می‌شود و در حالت رکورد با طول متغیر، چنین الزامی وجود ندارد.

۳ــ۲ــ۱ــ فایل[2] : مجموعه‌ای از رکوردها، تشکیل فایل را می‌دهند. به‌عنوان مثال : فایل تحصیلی دانش‌آموزان، محتوای فیلدهای رکوردهای مربوط به دانش‌آموزان را دارا می‌باشد.

۴ــ۲ــ۱ــ کلید[3] : به فیلد یا زیرمجموعه‌ای از فیلدها، که باعث یکتایی هریک از رکوردها شود، کلید گویند. به‌وسیلهٔ کلید می‌توان به یک رکورد مشخص دسترسی داشت. به‌عنوان مثال، به‌رکوردهای زیر توجه کنید :

ردیف	شماره دانش‌آموزی	نام	نام‌خانوادگی
۱	۱۰۰	علی	حسینی
۲	۱۱۰	حسن	احمدی
۳	۱۰۸	علی	علوی
۴	۹۴	رضا	علوی

۱ــ Record ۲ــ File ۳ــ Key

فیلد نام نمی‌تواند کلید باشد زیرا دو مقدار مشابه «علی» در رکوردهای ردیف ۱ و ۳ وجود دارد. همچنین فیلد نام‌خانوادگی نمی‌تواند کلید باشد زیرا مقادیر «علوی» در رکوردهای ردیف ۳ و ۴ وجود دارد. اما شماره دانش‌آموزی می‌تواند کلید باشد زیرا هیچ دو شماره دانش‌آموزی برابر وجود ندارد.

با توجه به رکوردهای این جدول، فیلدهای نام و نام‌خانوادگی نیز با هم می‌توانند کلید باشند. اگرچه در یک محیط عملیاتی مدرسهٔ واقعی، ممکن است دو دانش‌آموز نام و نام‌خانوادگی یکسان داشته باشند.

کنجکاوی

آیا ممکن است در یک سیستم ذخیره و بازیابی کلید وجود نداشته باشد؟

در انتخاب کلید موارد زیر قابل توجه است :

۱ــ کلید می‌تواند ساده یا مرکب باشد.

۲ــ ممکن است کلیدهای مختلفی اعم از ساده یا مرکب بتوان تشخیص داد (کلید کاندید) ولی یکی از آن‌ها را با توجه به ملاحظاتی به‌عنوان کلید اصلی[۱] درنظر می‌گیریم.

۳ــ در برخی موارد به‌جای استفاده از فیلدها و صفات مورد نیاز شناسایی شده، به‌صورت مجازی فیلدی با مقدار یکتا (صفت مجازی)، به‌عنوان کلید اضافه می‌شود. مانند شماره دانش‌آموزی. مجازی بودن یعنی یک فرد زمانی که در یک مدرسه ثبت‌نام می‌کند و دانش‌آموز آن مدرسه می‌شود، یک شمارهٔ دانش‌آموزی می‌گیرد و زمانی که فارغ‌التحصیل می‌شود دیگر این شماره را از دست می‌دهد.

۴ــ مقدار فیلدی که کلید درنظر گرفته می‌شود، تهی (خالی) نیست و طبق تعریف کلید، مقدار تکراری ندارد.

۳ـ۱ـ عملیات روی رکوردها

این عملیات عبارت است از :

۱ــ بازیابی رکوردها — Retrieve

۲ــ درج رکوردها — Insert

۱ــ Primary Key

۳ــ حذف رکوردها Delete

۴ــ به هنگام سازی رکوردها Update

۵ــ تغییر ساختار Restructure

۱ــ۳ــ۱ــ بازیابی رکوردها (Retrieve) : منظور از بازیابی به‌دست آوردن رکورد یا رکوردهایی است که دارای شرایط خاصی هستند. به‌عنوان مثال می‌خواهیم اطلاعات دانش‌آموزانی که معدل آن‌ها بیش از ۱۶ است را به‌دست آوریم. به این کار بازیابی گفته می‌شود. ممکن است نتیجهٔ بازیابی یک یا چند رکورد باشد.

به‌طور کلی برای دستیابی و بازیابی رکوردها دو روش وجود دارد :

۱ــ ترتیبی[۱]

۲ــ مستقیم[۲]

در روش ترتیبی، رکوردها از ابتدا یکی پس از دیگری بررسی می‌شوند تا «محتوای» موردنظر بازیافت شود. مثلاً فرض کنید می‌خواهیم منزل شخصی به‌نام علی علوی را در یک ساختمان چند طبقه پیدا کنیم. برای یافتن آپارتمان موردنظر باید مثلاً از طبقهٔ اول، زنگ واحدها را پشت سر هم بزنیم تا به منزل این شخص برسیم. این روش کُند و زمان‌بر است.

در روش مستقیم با توجه به محتوا، «آدرس» رکورد به‌دست می‌آید و رکورد موردنظر به‌طور مستقیم با آن آدرس بازیابی می‌شود. روش مستقیم به‌وسیلهٔ ساختارهای مختلف پیاده‌سازی می‌شود مثل شاخص[۳]. در مثال بالا اگر به دفتر سرایدار آپارتمان مراجعه کنیم، وی می‌گوید منزل این شخص مثلاً طبقهٔ چهارم، واحد ۲ است. ما می‌توانیم از آسانسور استفاده کنیم و سریع به منزل این شخص برسیم. حتی اگر آسانسور در دسترس نباشد از راه‌پله‌ها خود را به طبقه چهارم می‌رسانیم بدون آن که در طبقه‌ها وقت صرف جستجو کنیم[۴].

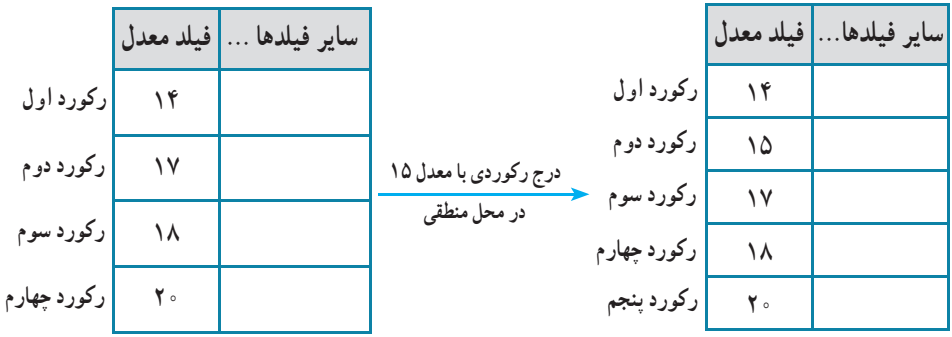
۲ــ۳ــ۱ــ درج رکوردها (Insert) : منظور، افزودن رکورد یا رکوردهایی است که تاکنون در سیستم وجود نداشته است. درج می‌تواند در انتهای فایل یا محل منطقی آن رکورد صورت گیرد. محل منطقی، محلی است که نظم فایل را حفظ می‌کند.

به‌عنوان مثال فرض کنید، فایلی به‌صورت زیر وجود دارد. اگر بخواهیم دانش‌آموزی با معدل ۱۵ را اضافه کنیم و نظم فایل که به‌صورت صعودی مقادیر معدل مرتب است حفظ شود، باید این رکورد

۱ــ Sequential ۲ــ Direct ۳ــ Index

۴ــ اگرچه از نظر زمانی نسبت به روش ترتیبی سریع‌تر است اما مصرف حافظه در آن بیشتر خواهد بود.

بین رکورد اول و دوم قرار گیرد. یعنی رکوردی بین اول و دوم اضافه کنیم و اطلاعات دانش‌آموز جدید را در آن ثبت کنیم.



شکل۲ــ۱ــ درج در محل منطقی (نظم صعودی فیلد معدل حفظ شده است)

۳ــ۳ــ۱ــ حذف رکوردها (Delete) : درصورتی که به رکورد یا رکوردهایی نیاز نداشته باشیم، آن را از فایل مربوطه حذف می‌کنیم.

۴ــ۳ــ۱ــ به هنگام‌سازی رکوردها (Update) : اصلاح مقادیر فیلد یا فیلدهایی از رکورد، در اصطلاح به‌هنگام‌سازی نام دارد. به‌عنوان مثال تغییر نمرهٔ یک درس دانش‌آموز به‌هنگام‌سازی محسوب می‌شود.

۵ــ۳ــ۱ــ تغییر ساختار(Restructure) : اگر به هر دلیل در ساختار رکوردها تغییر ایجاد شود مثلاً اندازهٔ فیلد تغییر کند یا فیلد جدیدی اضافه شود، تغییر ساختار صورت می‌گیرد.

۴ـ۱ـ سیستم فایلینگ (FS)

همان‌طورکه گفتیم سیستم فایلینگ، اولین نوع سیستم ذخیره و بازیابی اطلاعات است. در این سیستم، نرم‌افزارهای کاربردی، مدیریت ذخیره و بازیابی داده‌ها را نیز برعهده داشتند. در این سیستم برای کاربردهای مختلف، فایل‌های جداگانه‌ای طراحی و پیاده‌سازی می‌شوند. در سیستم فایلینگ انجام عملیات به‌وسیلهٔ برنامه‌سازی و پیاده‌سازی الگوریتم‌ها انجام می‌شود. زمانی که سیستم بزرگ و پیچیده می‌شود، هر تغییر جزیی در خواسته‌ها، منجر به کدنویسی می‌گردد.

به علت وجود فایل‌های مختلف، کنترل رعایت استاندارد در زمان پیاده‌سازی سخت است. به‌عنوان مثال فیلد نام‌خانوادگی در دو فایل ممکن است اندازه‌های مختلف داشته باشد. در سیستم فایلینگ، ممکن است «ناسازگاری داده‌ها» به‌وجود آید. درنظر بگیرید شخصی در زیر سیستم اول

خود را «علی علوی تهرانی» و در زیر سیستم دوم «علی علوی» معرفی کند. بدین‌ترتیب مثلاً در مقایسهٔ دو زیرسیستم، این دو مقدار یکسان نیستند و این به معنی ناسازگاری داده است. هم‌چنین احتمال بروز افزونگی[1] نیز وجود دارد. یعنی محتویات بعضی از فیلدها به‌صورت متعدد و تکراری ذخیره شوند. به‌طور مثال، مشخصات سکونت یک فرد هم در سیستم فایلینگ اول و هم در سیستم دوم ثبت گردد. نتیجهٔ افزونگی، اتلاف حافظه و مشکلات در انجام عملیات مبنایی است.

نکته

افزونگی، می‌تواند ناشی از طراحی نامناسب باشد که در سایر سیستم‌های ذخیره و بازیابی نظیر بانک اطلاعاتی نیز وجود دارد و منحصر به سیستم فایلینگ نیست.

برنامه‌سازی در این روش معمولاً با استفاده از یک زبان سطح بالا صورت می‌گیرد و معمولاً فایل‌های سیستم قابل استفاده به‌وسیلهٔ سایر سیستم‌های فایلینگ نیستند.

سیستم فایلینگ هنوز کاربرد دارد و به‌طورکلی معایب روش فایلینگ را می‌توان به‌صورت زیر خلاصه کرد :

۱ــ احتمال بروز افزونگی (تکرار در ذخیره‌سازی)

۲ــ حجم زیاد کدنویسی

۳ــ احتمال بالای ناسازگاری داده‌ها

۴ــ سخت بودن اعمال استاندارد در سیستم‌ها

۵ـ۱ـ پایگاه داده‌ها

سیستم مدیریت پایگاه داده، یکی از سیستم‌های توسعه یافتهٔ ذخیره و بازیابی اطلاعات است. اگرچه این مفهوم بسیار شناخته شده و رایج است، اما واقعیت آن است که تعریف کاملاً واحدی در متون معتبر برای آن ذکر نشده است.

در نگرش بانک اطلاعاتی، طراحی به‌صورت یکپارچه انجام می‌گیرد. از آن‌جایی که یک محیط عملیاتی می‌تواند دارای زیرمحیط‌های مختلفی باشد، این نگرش باعث می‌شود برخی معایب روش فایلینگ نظیر ناسازگاری داده‌ها و اعمال نشدن استانداردها رفع گردد، مثلاً برای فیلد نام در تمام محیط، مقدار

۱ــ Redundancy

یکسانی از حافظه اختصاص می‌یابد. حال آن‌که در روش فایلینگ به علت مجزا بودن فایل‌ها به سختی و با صرف هزینهٔ بالا می‌توان نگرش بانک اطلاعاتی داشت.

یکپارچه‌سازی و مدیریت متمرکز باعث جلوگیری از تکرار در ذخیره‌سازی (افزونگی) خواهد شد و درنتیجه از حافظه، استفاده بهتری صورت می‌گیرد. در مجموع می‌توان پایگاه داده را به‌صورت زیر تعریف کرد :

«مجموعه‌ای از داده‌ها به‌صورت یکپارچه با حداقل افزونگی، تحت کنترل یک سیستم متمرکز و در چارچوب یک مدل داده‌ای»

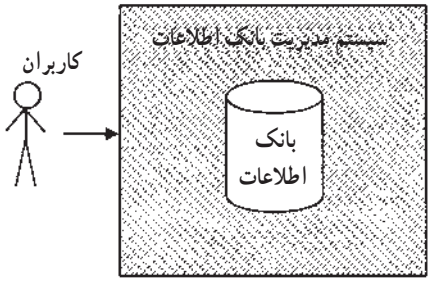
با ترکیب دست‌آوردهای شبکه‌ای، پایگاه داده می‌تواند به‌وسیلهٔ چند کاربر و همزمان استفاده شود.

کنجکاوی

تعریف‌های مختلف پایگاه داده را بررسی کنید.

۶ـ۱ ـ سیستم مدیریت پایگاه داده[۱]

اصلی‌ترین تفاوت روش پایگاه داده با روش‌های قبلی مثل سیستم فایلینگ، وجود حصاری به‌نام «سیستم مدیریت بانک اطلاعاتی» است. هرگونه دستیابی به داده‌ها باید از طریق آن انجام شود.



در DBMS، بانک اطلاعاتی و تمامی فایل‌های آن فقط و فقط در اختیار این نرم‌افزار قدرتمند هستند. کاربران درخواست خود را به این نرم‌افزار ارسال می‌کنند و در صورت تأیید، کار خواسته شده انجام می‌گیرد.

۱ــ۶ــ۱ــ معماری پایگاه داده : سؤال این است که چگونه می‌توان در یک سیستم بزرگ و پیچیدهٔ بانک اطلاعات، بین کاربران و داده‌های بسیار زیاد ذخیره شده روی رسانه ارتباط برقرار کرد و

۱ــ Database Management System (DBMS)

چگونه سیستم مدیریت پایگاه داده درخواست‌های کاربران برای عملیات پایگاه داده‌ای (مانند : بازیافت، درج، به‌هنگام‌سازی، حذف و غیره) را اداره و کنترل می‌کند.

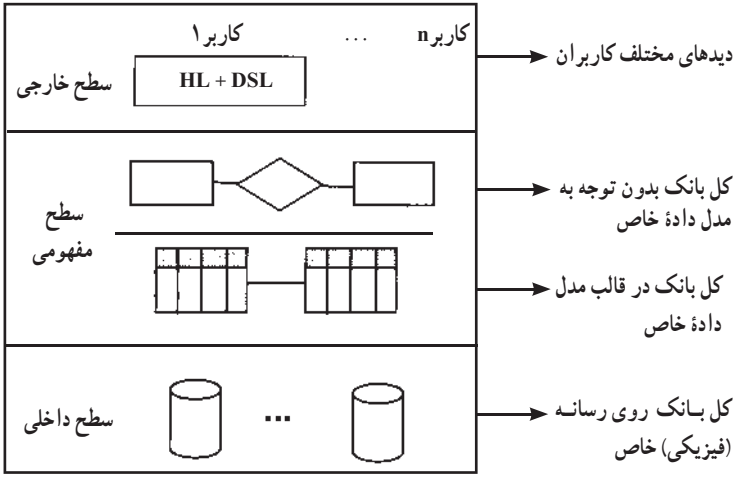
پاسخ این سؤال در معماری سه لایه‌ای است که به‌وسیلهٔ کمیته ANSI/SPARC[1] عرضه شده است. در ایجاد این معماری سطوح به‌گونه‌ای از هم مجزا شده‌اند که برای رسیدن به هدف با هم به‌خوبی مراوده و تعامل کنند و هم به اندازهٔ کافی مجزا و مستقل باشند که تغییر در یک سطح به سطوح دیگر انتقال پیدا نکند. این سه لایه عبارت است از :

۱ــ سطح خارجی[2]　　۲ــ سطح مفهومی[3]　　۳ــ سطح داخلی[4]

● ***سطح خارجی:*** همان دید کاربری است. بالاترین سطح انتزاع است. به این ترتیب که فقط بخشی از پایگاه داده که به مسایل کاربر و یا برنامهٔ کاربردی مربوط است را شامل می‌شود.

● ***سطح مفهومی:*** ساختار پایگاه داده است. تمام موجودیت‌ها[5] و ارتباطات بین آن‌ها را شامل می‌شود.

● ***سطح داخلی:*** جزییات ذخیره‌سازی است. پایین‌ترین سطح انتزاع است و به روش‌های فیزیکی ذخیره و بازیابی نزدیک است. چگونگی ذخیره‌سازی را نشان می‌دهد. ساختارهای داده را توضیح می‌دهد و هم‌چنین روش‌های دسترسی را نشان می‌دهد. سطح داخلی، سطح فایلینگ است.



ساختار سه لایه‌ای معماری پایگاه داده

۱ــ ANSI Standard Planning and Requirements Committee

۲ــ External Level　　۳ــ Conceptual Level　　۴ــ Internal Level

۵ــ Entity، مانند دانش‌آموز در محیط عملیاتی مدرسه

نکته

HL[1] : زبان برنامه‌نویسی غیربانکی است. مثل Visual Basic و...

DSL[2] : زبان برنامه‌نویسی بانکی یا به عبارتی زبان فرعی داده‌ای است. مثل SQL و...

۲ــ۶ــ۱ــ مدل داده‌ای : در تعریف پایگاه داده از «مدل داده‌ای» نام برده‌ایم. سیستم مدیریت بانک اطلاعاتی پنهان از دید کاربر، هرگونه دستیابی به داده‌ها را انجام می‌دهد. به این منظور در معماری پایگاه، مدل داده‌ای استفاده شده است. مدل‌های داده‌ای براساس ساختمان داده‌های مختلفی در بانک‌های اطلاعاتی استفاده می‌شوند که سه نوع متداول آن عبارت است از :

۱ــ رابطه‌ای Relational Data Structure

۲ــ سلسله مراتبی Hierarchical Data Structure[3]

۳ــ شبکه‌ای Network Data Structure[4]

پژوهش

در مورد مدل‌های داده‌ای دیگر پژوهش کنید.

همان‌طورکه دیدیم، سیستم مدیریت پایگاه داده‌ها، نرم‌افزاری است پیچیده که واسط بین کاربران و محیط فیزیکی ذخیرهٔ داده‌ها است. از جمله سیستم‌های مدیریت پایگاه اطلاعات می‌توان از : Oracle، My Sql، SQL Server، Access و... نام برد.

برخی از این سیستم‌های نرم‌افزاری از دیگری پیچیده‌تر، کامل‌تر و بنابراین گران‌تر هستند و برای ذخیره‌سازی داده‌های بسیار زیاد استفاده می‌شوند. به برخی از آن‌ها نیز غیر از مدیریت داده‌ها، بخش‌های دیگری اضافه شده است که واسط کاربری را ارایه می‌کنند. مثلاً در Oracle، بخش Development 2000 جدای از بخش مدیریت بانک عرضه می‌شود که طراح به کمک آن می‌تواند فرم‌ها و گزارش‌ها را ایجاد کند و آسان‌تر با محیط بانک ارتباط برقرار کند. در نرم‌افزار اکسس هم بخش‌هایی چون فرم‌ها، گزارش‌ها، ماجول‌ها و... برای این منظور ایجاد شده است که در بخش‌های بعدی بیشتر با آن آشنا می‌شویم.

۱ــ Host Language ۲ــ Data Sub Language

۳و۴ــ دربارهٔ مدل‌های سلسله مراتبی و شبکه‌ای در مقاطع بالاتر بیشتر خواهید آموخت.