



# مدرسان شریف

## فصل اول

### «مفاهیم پایگاه داده‌ها»

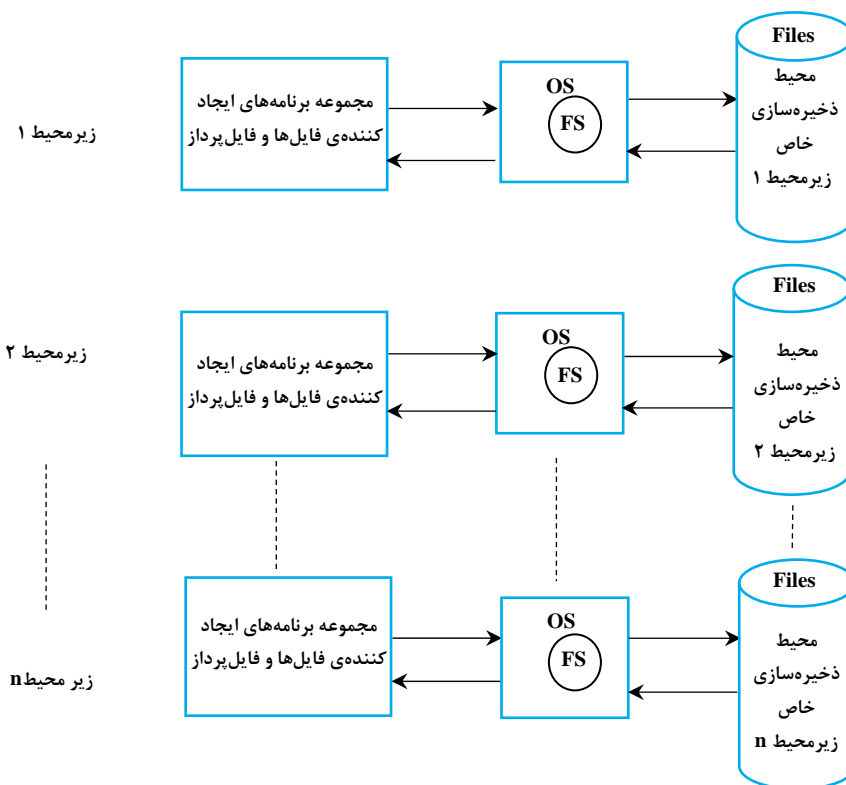
#### مقدمه

بشر تا به امروز برای ذخیره داده‌ها و اطلاعات و همچنین بازیابی آن‌ها، چالش‌های متفاوتی را پشت سر گذاشته و راهکارهای متفاوتی را ارائه نموده است. به طور کلی هدف مهمی این راهکارها افزایش کارایی در ذخیره و بازیابی اطلاعات است. یک سیستم ذخیره و بازیابی، سیستمی است که عملیات ذخیره و بازیابی اطلاعات را انجام می‌دهد و محیط عملیاتی (Operating Environment) محیطی است که برای آن سیستم ذخیره و بازیابی طراحی می‌شود مانند محیط عملیاتی دانشگاه، بیمارستان و ... روش‌های ذخیره و بازیابی را به طور کلی می‌توان به دو دسته طبقه بندی کرد:

۱- مشی فایلینگ (Filing)

۲- مشی پایگاهی

در مشی فایلینگ، هر محیط عملیاتی به زیر محیط‌های مجزا تقسیم می‌شود و به همین ترتیب اطلاعات فیزیکی زیر محیط‌ها نیز به طور مستقل از یکدیگر ذخیره می‌شود. ولی در مشی پایگاهی محیط عملیاتی به صورت واحد در نظر گرفته می‌شود و اطلاعات فیزیکی به صورت مجتمع وجود دارد. شکل (۱) معرف مشی فایلینگ است.



شکل ۱. نمودار مشی فایلینگ

- از ویژگی‌های مشی غیر بانکی می‌توان به موارد زیر اشاره کرد:
- زبان ایجاد سیستم و کار با سیستم یک زبان سطح بالا است.
  - تعدد سیستم‌ها و پراکندگی آن‌ها
  - وجود درجه‌ی بالایی از افزونگی (Redundancy)
  - عدم امکان اعمال استاندارد در محورهای مختلف
  - وجود ناسازگاری داده‌ها
  - سرعت پایین در بازیابی اطلاعات ترکیبی
  - هزینه‌ی بالای طراحی سیستم
  - عدم وجود ضوابط ایمنی کارا و مطمئن

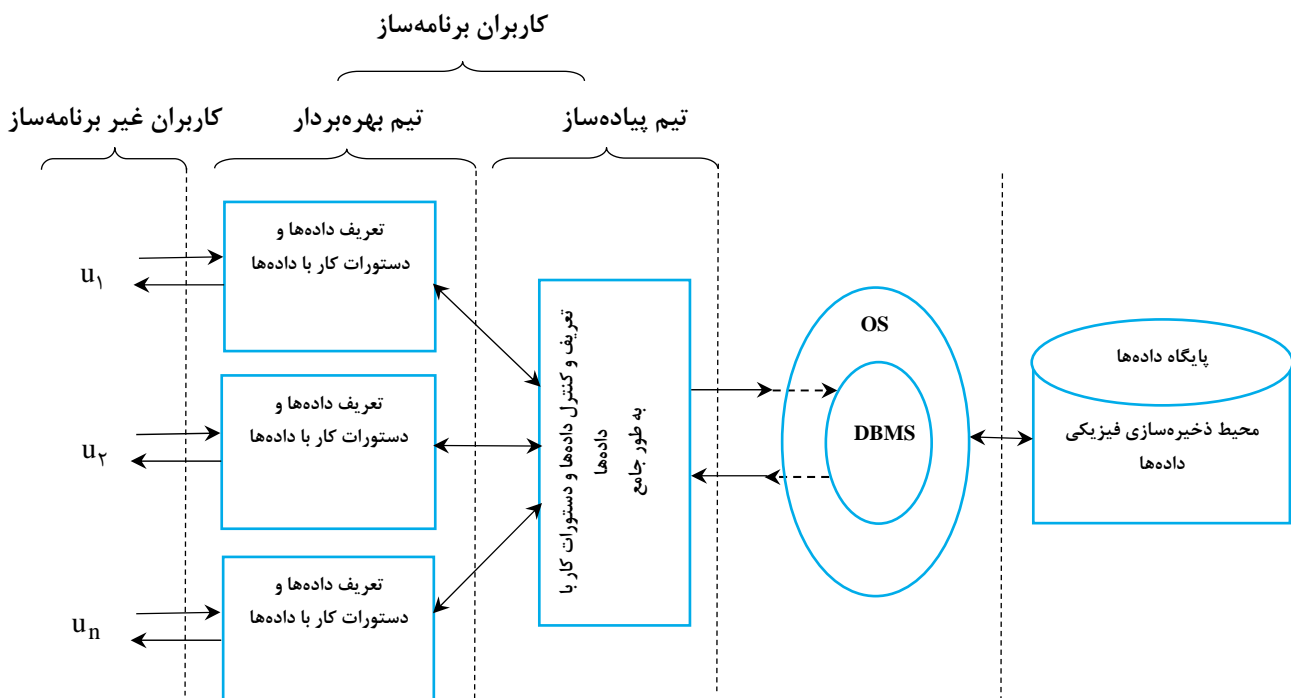
**مثال ۱:** کدام یک از گزینه‌ها از معایب سیستم‌های غیر پایگاهی است؟

- (۱) ناسازگاری داده‌ها  
 (۲) وابستگی برنامه‌های کاربردی به محیط ذخیره سازی  
 (۳) افزونگی داده‌ها  
 (۴) همه موارد

**پاسخ:** گزینه «۴» از دیگر معایب این سیستم می‌توان به عدم امکان اعمال استانداردها، عدم امکان قوانین جامعیتی، کاهش سرعت پرس‌وجو، کاهش سرعت سازماندهی مجدد و بروزرسانی نسبت به مشی پایگاهی و ... اشاره کرد.

برای رفع این معایب، مشی بانکی با ویژگی‌های مؤثر خود ایجاد شد که از مفهومی به نام پایگاه داده‌ها بهره می‌گیرد. در مشی پایگاهی مراحل کار به‌صورت زیر است:

- نیازهای اطلاعاتی اصلی برای همه‌ی قسمت‌های مورد توجه مدیریت آن سیستم توسط گروهی از تحلیل‌گران بررسی می‌شود.
  - داده‌های سیستم مدل‌سازی معنایی می‌شوند.
  - مشخصات سیستم یکپارچه‌ی مربوطه و وظایف آن تعیین می‌شود.
  - از سیستم مدیریت بانک اطلاعاتی به‌صورت متمرکز استفاده می‌شود.
  - طراحی پایگاه داده‌ها در سطوح و لایه‌های متفاوت انجام می‌شود.
  - محیطی مشترک و واحد برای ذخیره‌سازی مشخص می‌شود.
  - مجموعه‌ای از ابزارها برای بهره‌برداری از پایگاه داده‌ها طراحی می‌شود.
  - سیستم آزمایش نهایی می‌شود و بهره‌برداری اصلی صورت می‌گیرد.
- نمودار مشی پایگاهی به‌صورت شکل (۲) است:



شکل ۲: نمودار مشی پایگاهی

- ❖ **تعریف:** پایگاه داده‌ها (Database) مجموعه‌ای است از داده‌های ذخیره شده به طور مجتمع، حتی الامکان فاقد افزونگی، تعریف شده به صورت صوری، مبتنی بر یک ساختار، تحت کنترل یک سیستم مدیریت متمرکز، مورد استفاده‌ی یک یا چند کاربر، به طور همزمان و اشتراکی
- ❖ **تعریف:** سیستم مدیریت پایگاه داده‌ها (Database Management System) به مجموعه‌ای از داده‌های به هم مرتبط و برنامه‌هایی اطلاق می‌شود که قابلیت دسترسی به این داده‌ها را دارند و هدف اصلی آن ایجاد راهی برای ذخیره و بازیابی پایگاه داده‌ها به صورت کارا می‌باشد.

### 📌 مثال ۲: کدام گزینه تعریف درستی از پایگاه داده‌ها است؟

- (۱) مجموعه‌ای از داده‌های منطقاً مرتبط و توصیف این داده‌ها که برای پاسخ به نیازهای یک سیستم طراحی شده است.
- (۲) مجموعه‌ای از داده‌ها که بیانگر ارتباط و فعالیت میان چند سازمان است.
- (۳) مجموعه‌ای از داده‌های مانا که در سیستم کاربردی یک سازمان بکار گرفته می‌شود.
- (۴) همه موارد

✅ **پاسخ:** گزینه «۴» همه‌ی تعاریف ارائه شده، معرف پایگاه داده‌ها می‌باشند.

سیستم‌های پایگاه داده برای حجم بالایی از اطلاعات طراحی می‌شوند. این مدیریت شامل موارد زیر است:

- ۱- تعریف ساختارهایی برای ذخیره اطلاعات
  - ۲- ایجاد مکانیزم‌هایی برای تغییر در اطلاعات
  - ۳- تضمین امنیت اطلاعات ذخیره شده
- به طور کلی یک سیستم پایگاه داده‌ها موارد زیر که از معایب یک سیستم فایلینگ است را برطرف کرده است:
- ۱- **افزونگی داده‌ها و ناسازگاری داده‌ها (Data Inconsistency):** از آنجایی که برنامه‌نویسان متفاوت برای زیرسیستم‌های یک سیستم فایلینگ فعالیت می‌کنند بنابراین فایل‌های مختلفی با زبان‌های متفاوت ایجاد می‌شود. همچنین ممکن است اطلاعات مشابه در زیر سیستم‌های متفاوت تکرار شود. علاوه بر موارد گفته شده ممکن است اطلاعات مشابه در زیرسیستم‌های متفاوت ساختار ذخیره سازی یکسان نداشته باشند و یا تغییر این اطلاعات مشابه در یک زیرسیستم، در زیرسیستم‌های دیگر اعمال نشود.
  - ۲- **دشواری در دسترسی به داده‌ها:** در یک سیستم فایلینگ به دلیل مستقل بودن زیر سیستم‌ها، دسترسی به اطلاعات ترکیبی دشوار است و این‌گونه سیستم‌ها به داده‌ها اجازه نمی‌دهند تا به صورت مطلوب و کارا بازیابی شوند.
  - ۳- **انزوای داده‌ها (Data Isolation):** از آنجایی که در یک سیستم فایلینگ، داده‌ها در فایل‌های مستقل تقسیم می‌شوند و هر یک از فایل‌ها ممکن است دارای فرمت متفاوت باشند، نوشتن یک برنامه کاربردی برای بازیابی اطلاعات مطلوب دشوار است.
  - ۴- **مشکلات جامعیت (Integrity Problems):** منظور از جامعیت این است که ارزش داده‌های ذخیره شده باید دارای محدودیت‌های سازگاری داده‌ها باشد که اگر سیستم ذخیره و بازیابی از نوع فایلینگ باشد این مسئله پیچیده‌تر می‌شود. (در مورد جامعیت در بخش‌های بعد بیشتر توضیح داده می‌شود).
  - ۵- **مشکلات یکپارچگی (Atomicity Problems):** چالشی که در سیستم‌های مکانیکی و الکتریکی وجود دارد احتمال رخداد خرابی است و اینکه آیا در صورت ایجاد خرابی داده‌های موجود مجدداً در یک حالت مطلوب ذخیره‌سازی می‌شوند یا خیر. برقراری این امر در سیستم فایلینگ بسیار دشوار است.
  - ۶- **بی‌نظمی‌های موجود در دسترسی همروند (Concurrent Access Anomalies):** دسترسی به یک منبع مشترک به طور هم‌زمان، مشکلاتی را مانند نامعتبر بودن ارزش آن منبع به دنبال دارد که برای رفع این مشکل مدیریت دسترسی همروند نیاز است و این امر در سیستم‌های فایلینگ به سختی قابل پیاده‌سازی است.
  - ۷- **مشکلات امنیتی:** اینکه هر کاربری اجازه دسترسی به هر داده‌ای را نداشته باشد چالشی است که در سیستم‌های ذخیره و بازیابی وجود دارد و برقراری این امنیت در سیستم‌های فایلینگ به دلیل وجود زیر سیستم‌های مستقل بسیار دشوار است.

### 📌 مثال ۳: کدام یک از وظایف DBMS نیست؟

- (۱) کنترل امنیت پایگاه داده‌ها
  - (۲) کنترل سطح دسترسی
  - (۳) کنترل نرمال بودن پایگاه داده‌ها
  - (۴) کنترل همروندی تراکنش‌ها
- ✅ **پاسخ:** گزینه «۳» کنترل نرمال بودن پایگاه داده‌ها برعهده‌ی DBMS نیست و طراحان و مدیران پایگاه داده‌ها DBA هستند که بر نرمال‌سازی پایگاه داده‌ها نظارت دارند.

### 📌 مثال ۴: کدام گزینه از مزایای بانک اطلاعاتی نیست؟

- (۱) کاهش افزونگی داده‌ها
  - (۲) وابستگی داده‌ها
  - (۳) جلوگیری از ناسازگاری داده‌ها
  - (۴) اشتراک داده‌ها
- ✅ **پاسخ:** گزینه «۲» وابستگی داده‌ها در مفاهیم نرمال‌سازی و برای کاهش افزونگی و افزایش سطح نرمال اطلاعات در مدل رابطه‌ای کاربرد دارد. از مزایای دیگر بانک اطلاعاتی می‌توان به تعدد انواع کاربران و زبان‌ها، استقلال داده‌ای، امکانات سازماندهی مجدد، اعمال استانداردها، محرمانگی داده‌ها و ... اشاره کرد.



## عناصر اصلی پایگاه داده‌ها

چهار عنصر اصلی تشکیل‌دهنده محیط پایگاه داده‌ها عبارتند از:

- سخت‌افزار
- نرم‌افزار
- کاربر
- داده

### سخت‌افزار

به‌طور کلی سه دسته کلی سخت‌افزار در محیط پایگاه داده‌ها وجود دارد:

- ۱- **سخت‌افزار جهت ذخیره‌سازی داده‌ها:** در واقع رسانه‌هایی جهت ذخیره داده‌ها هستند که از ویژگی‌های آن‌ها می‌توان به سرعت بالا، ظرفیت بالا اشاره کرد.
- ۲- **سخت‌افزار جهت پردازش:** در واقع، سیستم کامپیوتری مربوط به پردازش پایگاه داده‌ها است که می‌تواند به‌طور عام یک کامپیوتر و یا به‌طور خاص برای کار با پایگاه داده‌ها طراحی شده باشد (DBM)
- ۳- **سخت‌افزار جهت برقراری ارتباط:** سخت‌افزارهایی جهت ایجاد یک ارتباط بین یک سیستم کامپیوتری و دستگاه‌های جانبی در یک سایت هستند و یا اینکه در سیستم پایگاه داده‌های نامتمرکز برای ارتباطات شبکه‌ای کاربرد دارند.

### نرم‌افزار

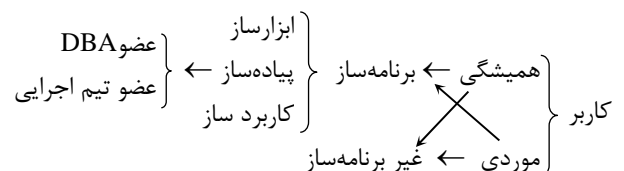
به‌طور کلی نرم‌افزارهای موجود در یک پایگاه داده‌ها عبارتند از:

- ۱- سیستم مدیریت پایگاه داده‌ها (DBMS)
- ۲- برنامه‌های کاربردی که می‌تواند به سه صورت نوشته شوند: الف) با زبان پایگاهی (ب) با زبان سطح بالا (ج) ترکیبی از هر دو
- ۳- رویه‌های ذخیره شده
- ۴- نرم‌افزارهای شبکه جهت ارتباطات شبکه‌ای

**نکته ۱:** DBMS در محیط سیستم عامل اجرا می‌شود و برنامه‌های کاربردی در محیط DBMS اجرا می‌شوند.

### کاربر

هر استفاده‌کننده از پایگاه داده‌ها یک کاربر محسوب می‌شود. دیاگرام زیر انواع کاربران پایگاه داده‌ها را نشان می‌دهد.



کاربر کاربردساز از Query Language یا ترکیبی از Host Language و Query Language برای برنامه‌سازی استفاده می‌کند. از طرف دیگر یک کاربر نهایی (End user) می‌تواند برنامه‌ساز یا غیربرنامه‌ساز باشد.

### داده

داده‌های ذخیره شده در پایگاه داده‌ها را اصطلاحاً داده‌های پایا می‌نامند و لزوماً مشابه با داده‌های ورودی و خروجی نیست. در یک پایگاه داده‌ها، برای ذخیره‌سازی داده‌ها باید مدل‌سازی معنایی انجام شود که اصطلاحاً طراحی ادراکی نامیده می‌شود. از روش‌های موسوم برای مدل‌سازی معنایی داده‌ها می‌توان به روش EER و UML و NIAM و OMT اشاره کرد.

### مثال ۵: مفهوم Information چیست؟

۲) داده‌های ذخیره شده در بانک اطلاعاتی

۱) داده‌های پردازش شده

۴) همه موارد

۳) داده‌هایی که روی آنها محاسباتی صورت گرفته است.

**پاسخ:** گزینه «۴» بطور کلی داده‌ای که پردازش شده است را اطلاعات می‌گویند و از آنجایی که داده‌های ذخیره شده در پایگاه داده‌ها، داده‌های پردازش شده هستند، عملاً در پایگاه داده‌ها مجموعه‌ای از اطلاعات ذخیره شده است.

## دید برای داده‌ها

از مهم‌ترین اهداف یک سیستم پایگاه داده‌ها، ایجاد یک دید انتزاعی از داده‌ها برای کاربران است که با این کار، سیستم پایگاه داده‌ها جزئیات ذخیره‌سازی و بازیابی را از دید کاربران پنهان می‌کند.

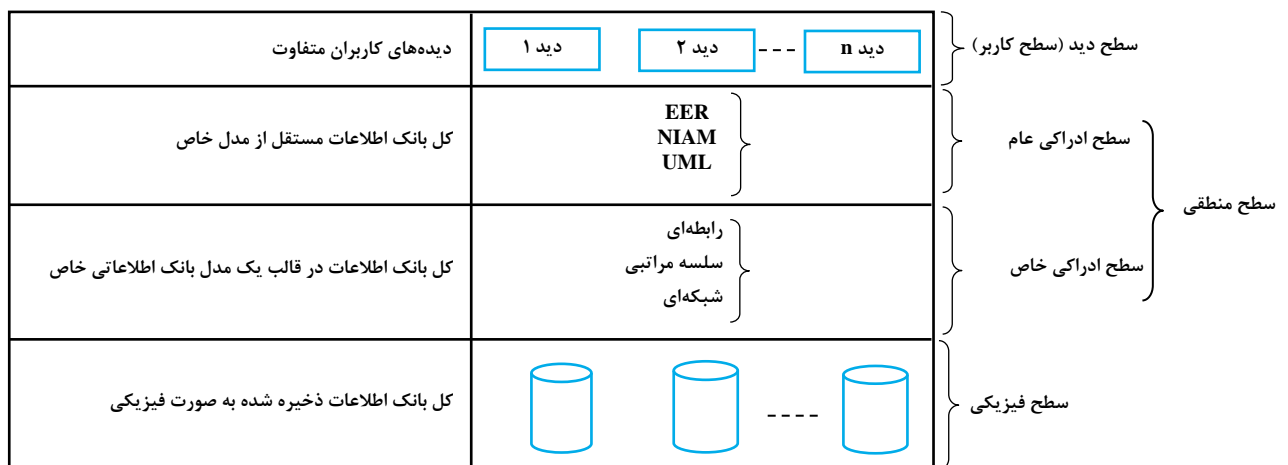
## انتزاع داده‌ها

به منظور ذخیره و بازیابی کارا در یک سیستم پایگاه داده‌ها، طراحان از ساختمان داده‌های پیچیده‌ای برای ذخیره‌سازی داده‌ها استفاده می‌کنند. از آنجایی که بیشتر کاربران سیستم پایگاه داده، اطلاعات تخصصی در این زمینه ندارند، طراحان این پیچیدگی را در چندین سطح انتزاع از دید کاربران پنهان می‌کنند تا ارتباط کاربران با سیستم ظاهر ساده‌ای داشته باشد. این سطوح انتزاع عبارتند از:

**سطح فیزیکی:** پایین‌ترین سطح انتزاع است که نحوه ذخیره‌سازی واقعی داده‌ها را نشان می‌دهد، این سطح شامل جزئیات ساختمان داده‌های پیچیده سطح پایین است که برای ذخیره‌سازی داده‌ها بکار رفته‌اند. نام دیگر این سطح، سطح داخلی است.

**سطح منطقی:** سطح بالای سطح فیزیکی است که بیان می‌کند چه داده‌هایی در پایگاه داده‌ها ذخیره شده‌اند و چه ارتباطی بین آن‌ها برقرار است. این سطح کل پایگاه داده‌ها را بصورت ساختارهای کوچک مرتبط نمایش می‌دهد. مدیران پایگاه داده‌ها مسئول کار با این سطح انتزاع هستند. نام دیگر این سطح، سطح ادراکی است که شامل دو زیر سطح ادراکی خاص و ادراکی عام است.

**سطح دید (سطح کاربران):** بالاترین سطح انتزاعی است که برای کاربران متفاوت، مختلف است و تنها شامل بخشی از کل پایگاه داده‌ها است. شکل (۳) بیانگر سطوح انتزاع داده‌ها در پایگاه داده‌ها می‌باشد.



شکل ۳: سطوح انتزاع داده در یک سیستم پایگاه داده‌ها

طراحی کلی یک پایگاه داده‌ها اصطلاحاً شمای پایگاه داده‌ها (Database Schema) نامیده می‌شود. متناظر با هر یک از سطوح انتزاع داده در بانک اطلاعات، یک شما به همان نام وجود دارد.

### کج مثال ۶: کدام یک از تصاویر زیر تنها در مرحله طراحی بانک اطلاعات مطرح است؟

- (۱) تصویر فیزیکی (۲) تصویر ادراکی عام (۳) تصویر خارجی (۴) تصویر ادراکی خاص

پاسخ: گزینه «۲» مدل‌سازی معنایی داده‌ها تنها در طراحی پایگاه داده‌ها کاربرد دارد و وجود خارجی ندارد و در سطح ادراکی عام مطرح می‌شود.

### کج مثال ۷: با توجه به سطح داخلی در معماری ANSI، DBMS از کدام مورد اطلاعی ندارد؟

- (۱) ساختار هر فایل (۲) پیاده‌سازی استراتژی دستیابی (Access method) (۳) توالی منطقی رکوردها (۴) پیوندهای منطقی بین فایل‌ها

پاسخ: گزینه «۲» سیستم مدیریت پایگاه داده‌ها از پیاده‌سازی منطقی فایل‌ها و ساختارشان اطلاع دارد ولی در نحوه پیاده‌سازی فیزیکی آنها و استراتژی‌های دستیابی به این اطلاعات بی‌اطلاع است.



## زبان‌های پایگاه داده‌ها

زبان‌های پایگاه داده‌ها به طور کلی به دو دسته کلی تقسیم می‌شوند:

۱- زبان میزبان Host Language (HL) یکی از زبان‌های سطح بالا یا اسمبلی است که به عنوان زبان پیاده‌سازی برنامه‌های کاربردی استفاده می‌شود مانند C, ++C و Java. زبان‌های میزبان فعلی به تنهایی برای تعریف و کار با پایگاه داده‌ها کافی نیستند به همین دلیل به زبان دیگری به عنوان زبان داده‌ای فرعی نیاز است.

۲- زبان داده‌ای فرعی (DSL) Data Sub Language

این زبان خود دارای دو دسته زبان می‌باشد:

الف - زبان تعریف داده‌ها (DDL (Data Definition Language

ب - زبان کار با داده‌ها (DML (Data Manipulation Language

به طور کلی زبان داده‌ای فرعی از لحاظ مکانیزم استفاده به دو دسته تقسیم می‌شود:

۱- مستقل: نیازی به HL ندارد مانند I. SQL

۲- ادغام شدنی (Embedded): دستورات DSL در یک HL استفاده می‌شود مانند E. SQL

در مورد ادغام DSL و HL نیز دو روش وجود دارد:

۱- ادغام صریح: در این حالت عین احکام DSL در متن برنامه HL آورده می‌شود.

۲- ادغام ضمنی: در این روش دستورات DSL صراحتاً در برنامه HL نوشته نمی‌شود و به صورت توابع با آنها برخورد می‌شود.

در مورد DSL باید نکات زیر رعایت شود:

۱- دستورات DSL باید در حوزه‌ی مفاهیم یک ساختمان داده تعریف شود.

۲- تعداد دستورات DSL باید حتی‌الامکان کم باشد.

۳- باید یادگیری آن ساده و در استفاده آسان باشد.

۴- DSL باید اکمال ساختاری (Structural Completeness) داشته باشد. برای مثال اگر ساختمان داده جدول است، DSL باید بتواند هرگونه جدول ممکن را از پایگاه استخراج کند.

۵- باید در DSL اصل «وحدت دستورات» رعایت شود. یعنی برای هر عمل منطقاً واحد، دستور واحد موجود باشد.

**مثال ۸:** کدام گزینه نادرست است؟

(۱) دستورات DSL باید حتی‌الامکان کم باشد.

(۲) در DSL می‌توان از زبان‌های سطح بالا مانند C استفاده کرد.

(۳) I. SQL نمونه‌ای از DSL‌های ادغام شدنی است.

(۴) ترکیبی از زبان‌های سطح بالا و زبان‌های توصیفی، زبان کار با پایگاه داده‌ها را ایجاد می‌کند.

پاسخ: گزینه «۳» از خصوصیات دستورات DSL سادگی در فهم و کوتاهی آن است این زبان‌ها به دو صورت صریح و ادغام شدنی استفاده می‌شود. در روش ادغام شدنی ترکیبی از یک زبان توصیفی مانند SQL و یک HL مانند C و ++C وجود دارد. I. SQL از نوع DSL‌ها به روش صریح است.

### زبان تعریف داده‌ها (DDL)

شمای پایگاه داده توسط مجموعه دستوراتی از زبان خاص ایجاد می‌شود که اصطلاحاً به آن زبان تعریف داده‌ها گفته می‌شود. زبان تعریف داده‌ها علاوه بر ایجاد شمای پایگاه داده‌ها، مجموعه‌ی خاصی از جداول که لغت‌نامه‌ی داده‌ها نامیده می‌شود را نیز بروز رسانی می‌کند.

یک لغت‌نامه‌ی داده شامل «فرا داده» Meta Data است (فرا داده، داده‌ای است در مورد داده‌ی دیگر). شمای یک جدول مثالی از فراداده است. همچنین حق دستیابی کاربران مختلف نیز یک فراداده است. داده‌های ذخیره شده در پایگاه داده‌ها دارای محدودیت‌های سازگاری می‌باشند که زبان تعریف داده‌ها قابلیت تعریف این محدودیت‌های سازگاری داده را نیز دارد.

### زبان کار با داده‌ها (DML)

زبان کار با داده‌ها زبانی است که به کاربران این امکان را می‌دهد که به داده‌های سازماندهی شده در مدل‌های داده‌ای دسترسی داشته و آن‌ها را تغییر دهند و شامل موارد زیر است:

- ۱- بازیابی اطلاعات ذخیره شده در پایگاه داده‌ها
- ۲- درج اطلاعات جدید در پایگاه داده‌ها
- ۳- حذف اطلاعات از پایگاه داده‌ها
- ۴- تغییر اطلاعات ذخیره شده در پایگاه داده‌ها

به طور کلی دو نوع زبان کار با داده‌ها وجود دارد:

۱- زبان کار با داده‌های مبتنی بر رویه (Procedural DML): این نوع DML نیاز به یک کاربر دارد تا مشخص کند چه اطلاعاتی نیاز است و چگونه باید بازیابی شود.

۲- زبان کار با داده‌های توصیفی (Declarative DML): تنها برای اینکه تعیین شود چه اطلاعاتی باید بازیابی شود به کاربر نیاز دارد و در مورد چگونگی بازیابی اطلاعات نیازی به کاربر نیست. زبان‌های توصیفی نسبت به زبان‌های رویه‌ای دارای ویژگی سهولت در یادگیری و استفاده هستند.

**نکته ۲:** برنامه‌های کاربردی، برنامه‌هایی هستند که برای ارتباط با پایگاه داده‌ها از آن‌ها استفاده می‌شود و به یکی از زبان‌های HL (C, C++, Java, Cobol, ...) نوشته می‌شوند.

**نکته ۳:** به منظور دسترسی به پایگاه داده‌ها دستورات DML باید در داخل زبان HL اجرا شوند که این امر به دو صورت انجام می‌شود:

الف - توسط یک واسط برنامه کاربردی که برای ارسال DDL و DML به پایگاه داده‌ها و بازیابی نتایج استفاده می‌شود.

ب - گسترش HL و ادغام DML در داخل آن که این امر نیاز به یک پیش کامپایلر را برای تبدیل اولیه دستورات DML به فرم زبان HL، تأیید می‌کند.

**مثال ۹:** کدام گزینه صحیح است؟

۱) DBMS از نظر ماهیت زبان داده‌ای فرعی به دو دسته رویه‌ای و غیر رویه‌ای تقسیم می‌شوند.

۲) سیستم فایل تمام DBMSها وابسته به سیستم عامل است.

۳) ساختار داده‌ای مورد استفاده در DBMSها شامل دو دسته مدل رابطه‌ای و شی‌گرا هستند.

۴) DBMSها همه منظوره طراحی می‌شوند و DBMS تک‌منظوره وجود ندارد.

پاسخ: گزینه «۱» در مورد گزینه (۲) الزامی وجود ندارد که همه سیستم فایل‌ها وابسته به سیستم عامل باشند. در مورد گزینه (۳) مدل‌های داده‌ای دیگری نیز مانند شبکه‌ای و استنتاجی و سلسله‌مراتبی نیز وجود دارد. گزینه (۴) نیز نادرست است زیرا DBMSهای تک‌منظوره نیز برای کاربردهای خاص طراحی می‌شوند.

## کاربران و مدیران پایگاه داده‌ها

به طور کلی افرادی که با پایگاه داده‌ها کار می‌کنند را می‌توان به دو دسته تقسیم کرد:

۱- کاربران پایگاه داده‌ها      ۲- مدیران پایگاه داده‌ها (DBA)

### کاربران پایگاه داده‌ها

به طور کلی بر اساس نحوه‌ی کار با پایگاه داده‌ها چهار نوع کاربر وجود دارد:

۱- کاربر ساده: کاربران ساده، کاربرانی هستند که تنها با فراخوانی یک برنامه کاربردی از سیستم با پایگاه داده‌ها ارتباط برقرار می‌کنند.

۲- برنامه‌نویسان کاربردی: افراد حرفه‌ای در کامپیوتر که قابلیت برنامه‌نویسی کاربردی دارند و می‌توانند واسط کاربری مورد نیاز خود را خودشان طراحی کنند، این طراحی می‌تواند توسط ابزارهای RAD نیز بدون برنامه‌نویسی انجام شود.

۳- کاربران پیچیده: این کاربران با سیستم، بدون برنامه‌نویسی ارتباط برقرار می‌کنند و درخواست‌های خود را در قالب زبان پرس‌وجو (Query Language) مطرح می‌کنند.

۴- کاربران خاص: کاربران پیچیده‌ای هستند که برنامه کاربردی خاص پایگاه داده‌ها را می‌نویسند که در چارچوب پردازش داده‌ی سنتی قرار ندارد مانند سیستم‌های خبره، پایگاه دانش و ...

### مدیران پایگاه داده

یکی از مهم‌ترین دلایل استفاده از DBMS، کنترل مرکزی داده‌ها و برنامه‌هایی است که با این داده‌ها کار می‌کنند. شخص یا اشخاصی که وظیفه کنترل سیستم پایگاه داده‌ها را برعهده دارد مدیر پایگاه داده‌ها (DBA) Data Base Administrator نامیده می‌شود. از وظایف DBA می‌توان به موارد زیر اشاره کرد:

۱- تعریف شما: یک DBA با اجرای دستورات DDL شما می‌تواند پایگاه داده‌ها را ایجاد می‌کند.

۲- تعریف ساختمان داده‌های ذخیره‌سازی و متدهای دسترسی

۳- ایجاد تغییرات در سازمان فیزیکی شما پایگاه داده‌ها

۴- تعیین حق دسترسی به داده‌های ذخیره شده در پایگاه داده‌ها

۵- تعریف سیاست‌های مربوط به ترمیم و پشتیبانی

۶- نگهداری و پشتیبانی همیشگی: مانند گرفتن فایل پشتیبان، بررسی کارهای در حال اجرا، کنترل فضای خالی دیسک و ...

۷- مسئولیت مستقیم برنامه‌نویسان پایگاه داده‌ها

۸- مسئولیت وضع استانداردها



### کلمه مثال ۱۰: مفهوم File System چیست؟

(۱) انباره‌ای برای یک مجموعه از فایل‌های داده‌ای

(۲) به ساختار نامگذاری و ذخیره‌سازی فایل‌ها در یک سیستم عامل گفته می‌شود.

(۳) همان DBMS است.

(۴) همه موارد

پاسخ: گزینه «۲» فایل سیستم بخشی است که ممکن است وابسته به سیستم عامل یا مستقل از آن فعالیت کند و جزیی از DBMS نیز است که نحوه ذخیره‌سازی فایل‌ها را کنترل می‌کند.

### تراکنش (Transaction)

به طور کلی هر برنامه‌ای که توسط کاربران در پایگاه داده‌ها اجرا می‌شود یک تراکنش است. یک تراکنش همواره توسط DBMS کنترل می‌شود و ممکن است اجرای آن به دلایلی به تعویق بیفتد یا حتی تراکنش از اجرا ساقط شود. دلیل اصلی این امر برقراری صحت و جامعیت داده‌های پایگاه داده‌ها است. زیرا در یک پایگاه داده‌ها رکن اصلی، داده می‌باشد و حفظ صحت و جامعیت داده مهم‌ترین هدف یک DBMS است.

در سال ۱۹۸۱ جیم - گری ثابت کرد برای برقراری صحت و جامعیت پایگاه داده‌ها چهار کنترل روی همه تراکنش‌ها باید انجام شود که عبارتند از:

#### الف - یکپارچگی (Atomicity)

این خاصیت می‌گوید یا همه دستورهای یک تراکنش به طور کامل اجرا می‌شود و یا هیچکدام اجرا نمی‌شود.

#### ب - سازگاری (Consistency)

این خاصیت می‌گوید که هر تراکنش باید تمام قوانین جامعیت بانک اطلاعات را رعایت کند و اینکه فرض شود یک تراکنش یک برنامه صحیح است. با برقراری این خاصیت، هر تراکنش اگر به تنهایی اجرا شود، پایگاه داده‌ها را از یک حالت صحیح به حالت صحیح دیگری می‌برد.

#### ج - انزوا (Isolation)

این خاصیت مربوط به تراکنش‌های همروند است و می‌گوید که اثر تراکنش‌های همروند نباید بر روی یکدیگر مخرب باشد. این خاصیت توسط بخشی از DBMS به نام واحد کنترل همروندی فراهم می‌شود.

#### د - مانایی (Durability)

این خاصیت می‌گوید، تراکنش‌هایی که پایان موفق داشته‌اند (Commit) اثرشان همیشگی و مانا است و هرگز بدون دلیل از بین نمی‌روند. پایان ناموفق یک تراکنش اصطلاحاً Abort نامیده می‌شود.

نکته ۴: دو خاصیت مانایی و یکپارچگی توسط بخشی به نام مدیریت بازگرد فراهم می‌شود. همچنین چهار خاصیت تراکنش به خواص ACID نیز شهرت دارد.

### کلمه مثال ۱۱: در پایگاه داده‌ی دانشگاه، قاعده‌ی ذکر شده، بیانگر کدام خاصیت ACID در تراکنش‌ها است؟ «ظرفیت هر گروه نباید بیش از ۲۰ نفر باشد.»

(۴) همخوانی

(۳) مانایی

(۲) انزوا

(۱) یکپارچگی

پاسخ: گزینه «۴» اعمال قواعد جامعیت از جمله قاعده‌ی سازگاری داده‌ها، باعث برقراری صحت و جامعیت در تراکنش‌ها می‌شود و قاعده‌ی ذکر شده، همخوانی داده‌ها را بیان می‌کند.

### کلمه مثال ۱۲: کدام گزینه صحیح نیست؟

(۱) هر برنامه‌ای که در محیط بانک اطلاعات اجرا شود یک تراکنش است.

(۲) تراکنش‌ها در محیط بانک اطلاعات مانا نیستند.

(۳) در محیط بانک اطلاعات Roll Back به معنای اجرای ناموفق تراکنش است.

(۴) اجرای موفقیت‌آمیز یک تراکنش Commit نامیده می‌شود.

پاسخ: گزینه «۳» به اجرای ناموفق یک تراکنش اصطلاحاً Abort گفته می‌شود. نکته مهم در این تست گزینه (۲) است. علی‌رغم اینکه به ظاهر این گزاره نادرست است ولی باید توجه داشت که در حالت کلی تراکنش‌ها مانا نیستند ولی اثر تراکنش‌ها خاصیت پایایی دارد.



## استقلال داده‌ای

استقلال داده‌ای (Data Independence) مصونیت برنامه‌های کاربردی و دید کاربران در سطح خارجی در برابر تغییرات اعمال شده در سطوح پایین‌تر پایگاه داده‌ها است. به عبارت دیگر عدم وابستگی سطح خارجی به دیگر سطوح را استقلال داده‌ای می‌گویند. استقلال داده‌ای از مهم‌ترین مزایای تکنولوژی پایگاهی است.

استقلال داده‌ای بر دو نوع است:

### ۱- استقلال داده‌ای فیزیکی (Physical Data Independence)

### ۲- استقلال داده‌ای منطقی (Logical Data Independence)

استقلال داده‌ای فیزیکی، مصونیت برنامه‌های کاربردی و دید کاربران در برابر تغییرات سطح فیزیکی پایگاه داده‌ها است. زیرا کاربران در یک سطح کاملاً انتزاعی نسبت به فایلینگ پایگاه داده‌ها هستند. این تغییرات در سطح فیلد، رکورد، بلاک، فایل و ... قابل انجام است. استقلال داده‌ای فیزیکی در سیستم‌های رابطه‌ای جدید به طور کامل تأمین می‌شود زیرا کاربران در سطح خارجی، در یک محیط انتزاعی و بدون تماس با سطح فایلینگ پایگاه داده‌ها قرار دارند و همچنین بین سطح خارجی و سطح داخلی، یک سطح انتزاعی دیگر به نام سطح ادراکی وجود دارد.

استقلال داده‌ای منطقی، مصونیت برنامه‌های کاربردی و دید کاربران در برابر تغییرات انجام شده در سطح ادراکی پایگاه داده‌ها است. این تغییرات می‌تواند به دلایل زیر باشد:

- ۱- رشد پایگاه داده‌ها در سطح ادراکی که به دلیل مطرح شدن نیازهای اطلاعاتی جدید مانند صفت خاصی جدید، موجودیت جدید، ارتباط جدید اتفاق افتاده است.
- ۲- سازماندهی مجدد سطح ادراکی که به دلایلی از قبیل افزایش ایمنی پایگاه داده‌ها، تأمین محیط کارآتر برای ذخیره‌سازی داده‌ها، طراحی بهتر به منظور تأمین رفتار بهتر برای DBMS و ... انجام می‌شود.

تأمین استقلال داده‌ای منطقی در سیستم‌های رابطه‌ای امروزی به طور صددرصد انجام نمی‌پذیرد. می‌توان گفت رشد پایگاه داده‌ها در تأمین استقلال داده‌ای منطقی مشکلی ایجاد نمی‌کند و سازماندهی مجدد پایگاه داده‌ها (مانند تجزیه یک رابطه به چند رابطه‌ای کوچکتر) می‌تواند باعث نقض استقلال داده‌ای منطقی شود.

### کج مثال ۱۳: کدام گزینه مفهوم استقلال داده‌ای را کامل‌تر بیان می‌کند؟

- ۱) تغییرات پایگاه داده‌ها در سطح منطقی به سطح کاربران منتقل نمی‌شود.
- ۲) تغییرات پایگاه داده‌ها در سطح فیزیکی به سطح کاربران منتقل می‌شود.
- ۳) تغییرات پایگاه داده‌ها در سطح فیزیکی به سطح کاربران منتقل نمی‌شود.
- ۴) (۱) و (۳)

پاسخ: گزینه «۴» استقلال داده‌ای به دو صورت منطقی و فیزیکی تعریف می‌شود. استقلال داده‌ای منطقی در واقع وجود یک انتزاع بین سطح مفهومی و سطح کاربران را نشان می‌دهد و استقلال داده‌ای فیزیکی این انتزاع را میان سطح فیزیکی و کاربران عنوان می‌کند.

### کج مثال ۱۴: مزیت تعریف دید بر روی دید (View on View) چیست؟

- ۱) تأمین کننده‌ی جامعیت پایگاه داده‌ها
- ۲) کاهش سربار در DBMS
- ۳) تأمین کننده‌ی استقلال داده‌ای منطقی
- ۴) به اشتراک گذاشتن داده‌ها

پاسخ: گزینه «۳» هدف اصلی از تعریف دید (View) فراهم کردن استقلال داده‌ای منطقی است تا یک انتزاع میان سطح مفهومی و سطح کاربران حاصل شود.

## ساختار سیستم مدیریت پایگاه داده‌ها (DBMS)

بخش عملیاتی یک DBMS می‌تواند به دو زیر بخش اصلی تقسیم شود که عبارت است از مدیر ذخیره‌سازی (Storage Manager) و پردازشگر پرس و جو (Query Processor)

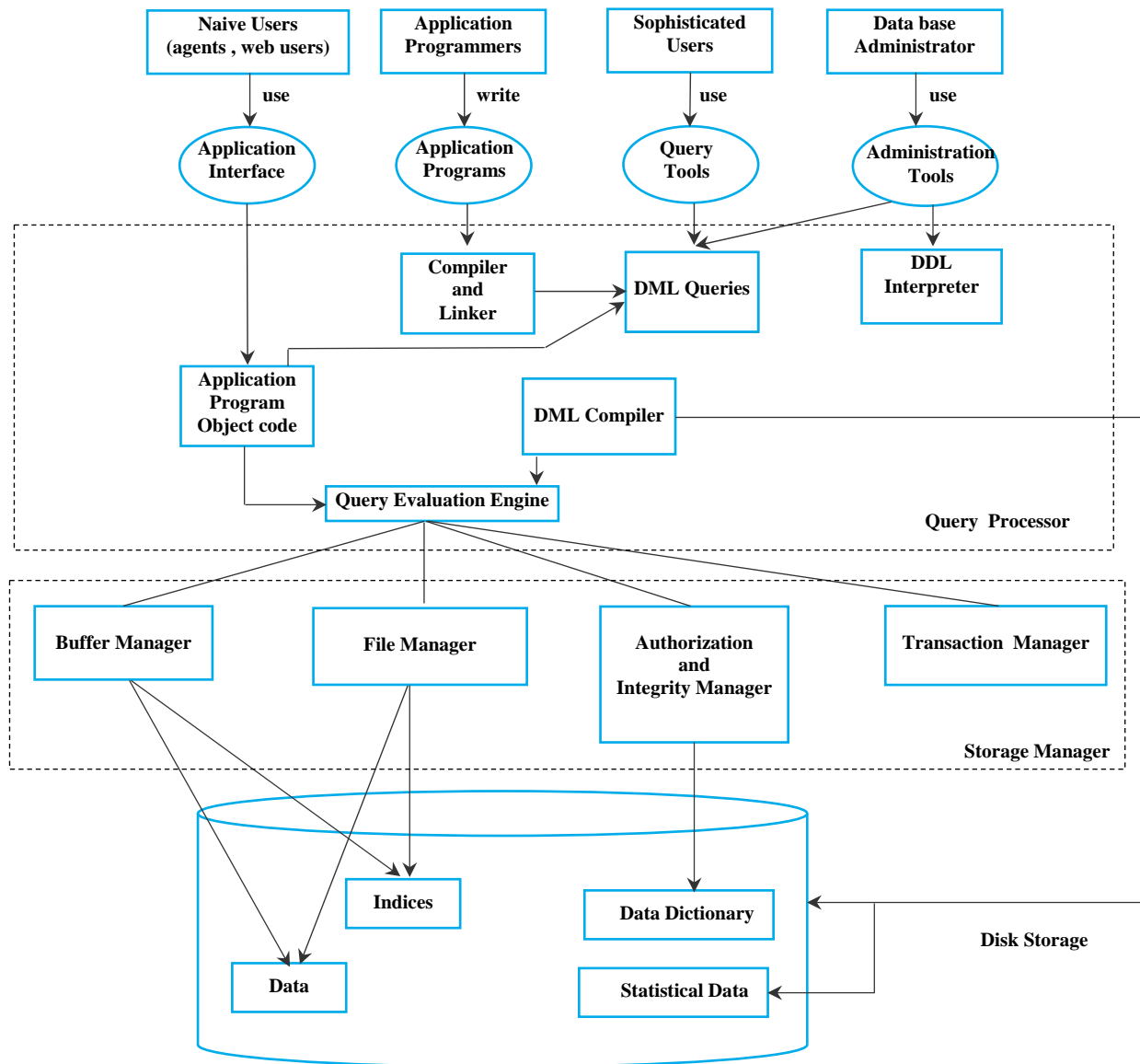
مدیر ذخیره‌سازی از این نظر اهمیت دارد که فضای مورد نیاز برای پایگاه داده‌ها بسیار زیاد است و تنها می‌توان این فضا را بر روی دیسک فراهم کرد و اطلاعات لازم باید به حافظه‌ی اصلی منتقل شوند. بنابراین از آنجایی که سرعت انتقال اطلاعات میان دیسک و حافظه‌ی اصلی کند است، مدیر ذخیره‌سازی بهترین گزینه برای کاهش این انتقال است.

یک مدیر ذخیره‌سازی مسئول برقراری ارتباط با مدیریت فایل سیستم است. همچنین وظیفه‌ی ترجمه‌ی DML به دستورات سطح پایین سیستمی را نیز بر عهده دارد. بنابراین یک مدیر ذخیره‌سازی، ماجولی (پیمانه‌ای) است که وظیفه‌ی ذخیره‌سازی، بازیابی و بروزرسانی داده‌های پایگاه داده‌ها را برعهده دارد. زیربخش‌های مدیر ذخیره‌سازی داده‌ها عبارتند از:

- مدیریت جامعیت و حق دسترسی: به قواعد جامعیت داده‌ها و حق دستیابی کاربران برای دسترسی به داده‌ها می‌پردازد.
  - مدیریت تراکنش: چهار خاصیت ACID را فراهم می‌کند.
  - مدیریت فایل: وظیفه بررسی فضای تخصیص داده شده به داده‌ها و ساختارهای داده‌ای به کار رفته در ذخیره سازی داده‌ها را برعهده دارد.
  - مدیریت بافر: وظیفه واکنشی داده‌ها از دیسک به حافظه اصلی را برعهده دارد.
- همچنین مدیر ذخیره‌سازی داده‌ها، ساختمان داده‌های متعددی را به عنوان سطح فیزیکی پایگاه داده‌ها پیاده‌سازی می‌کند که عبارتند از:
- فایل‌های داده‌ای
  - لغت‌نامه‌ی داده‌ها
  - شاخص‌ها (indices)

پردازشگر پرس‌وجو نیز از این جهت اهمیت دارد که باعث می‌شود DBMS به داده‌های مورد نیاز سریع‌تر و آسان‌تر دسترسی پیدا کند. همچنین این بخش وظیفه ترجمه‌ی بروز رسانی‌ها و پرس‌وجوهای نوشته شده به زبان توصیفی را به مجموعه دستورات منطقی در سطح فیزیکی برعهده دارد. زیربخش‌های این واحد عبارتند از:

- **مفسر DDL**: وظیفه ترجمه‌ی دستورات DDL و ذخیره آن‌ها در لغت‌نامه‌ی داده‌ها را برعهده دارد.
  - **کامپایلر DML**: وظیفه ترجمه دستورات DML به مجموعه‌ای از دستورات سطح پایین و قابل فهم برای موتور پرس‌وجو را برعهده دارد.
  - **موتور محاسبه‌ی پرس‌وجو**: وظیفه اجرای دستورات سطح پایین تولید شده توسط کامپایلر DML را دارد.
- ارتباط کامل این واحدها در یک DBMS در شکل (۴) نشان داده شده است.



شکل ۴. ساختار DBMS