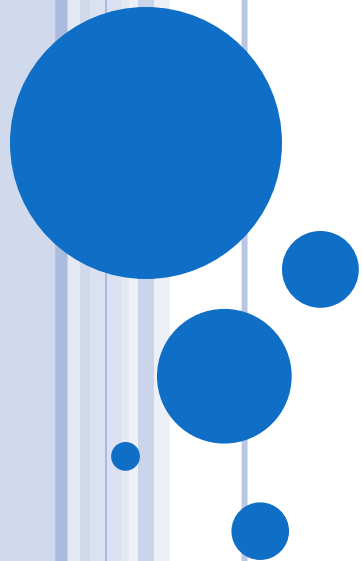


پایگاه داده ها

Database - SQL



فایل های اطلاعاتی که به نوعی به هم مرتبط هستند، تشکیل یک بانک اطلاعاتی را می دهند. فایل شامل مجموعه ای از رکوردها می باشد و رکورد مجموعه ای از فیلدهای به هم مرتبط است و فیلد کوچکترین جزء یک بانک اطلاعاتی می باشد. مثلا در بانک اطلاعاتی دانشگاه چندین فایل وجود دارد مانند فایل دانشجویان که شامل چندین رکورد است، هر رکورد شامل اطلاعات یک دانشجو می باشد که از چند فیلد مانند شماره دانشجویی، نام، آدرس، معدل و... تشکیل شده است.

داده: نمایش پدیده ها و مفاهیم به صورت صوری و مناسب برای برقراری ارتباط یا پردازش.

اطلاع: داده پردازش شده می باشد.

شناخت: نمایش نمادین جنبه هایی از بخشی از جهان واقع می باشد. به عبارتی نوعی اطلاع سطح بالاتر است.

پایگاه داده ها: مجموعه ای از داده های ذخیره شده و پایا به صورت مجتمع و بهم مرتبط، با کمترین افزونگی، تحت مدیریت یک سیستم کنترل متمرکز، مورد استفاده یک یا چند کاربر به صورت همزمان و اشتراکی.

۱- روش فایلینگ (ناپایگاهی)

۲- روش پایگاهی

۱- روش فایلینگ (ناپایگاهی)

در روش فایلینگ (سنتی)، نیازهای اطلاعاتی و پردازشی هر قسمت از محیط برآورده می شوند. مراحل اولیه طراحی و تولید برای هر قسمت به طور کلاسیک انجام شده و بعد از طراحی، مشخصات هر سیستم همراه با وظایف آنها مشخص می شود. در این روش، برای ایجاد محیط ذخیره سازی اطلاعات از یک سیستم فایل (FS) و برای برنامه سازی از یک زبان سطح بالا استفاده می شود و در نهایت برای هر قسمت، یک سیستم کاربردی ایجاد می شود.

معایب روش فایلینگ

- ۱- عدم وجود محیط مجتمع ذخیره سازی
- ۲- عدم وجود سیستم کنترل متمرکز
- ۳- عدم وجود ضوابط ایمنی کارا
- ۴- عدم امکان اشتراکی شدن داده ها
- ۵- تکرار در ذخیره سازی اطلاعات
- ۶- مصرف نامناسب امکانات سخت افزاری و نرم افزاری
- ۷- وابسته بودن برنامه های کاربردی به محیط ذخیره سازی داده ها
- ۸- حجم زیاد برنامه سازی

۲- روش پایگاهی

در این روش نیازهای اطلاعاتی تمامی قسمتها مورد مطالعه قرار می گیرد تا بتوان یک سیستم یکپارچه (integrated) طراحی کرد. داده های سازمان مدلسازی معنایی (SDM) می شوند و مشخصات سیستم یکپارچه تعیین می شود. برای سیستم مدیریت متمرکز از یک یا چند DBMS استفاده می شود. طراحی پایگاه داده ها در سطوح لازم انجام می شود و کاربران هر قسمت، پایگاه داده های خود را تعریف می کنند و با آن کار می کنند.

در واقع در روش پایگاهی یک محیط ذخیره سازی واحد، مجتمع و اشتراکی، تحت کنترل متمرکز وجود دارد که کاربران براساس نیاز خاص خود، پایگاه خود را تعریف کرده و هر کاربر تصور می کند که پایگاه خود را دارد.

روش های ایجاد سیستم های کاربردی

- ✓ کاربران در روش پایگاهی بطور همزمان از سیستم استفاده می کنند.
- ✓ در روش پایگاهی نسبت به داده های ذخیره شده، تنوع و کثرت دید وجود دارد.
- ✓ در روش پایگاهی نسبت به روش فایلینگ، حجم برنامه ها کمتر و برنامه سازی آسانتر است.
- ✓ پایگاه داده ها بر حسب تعداد رکوردهای آن، به دسته های کوچک، متوسط، بزرگ (LDB) و خیلی بزرگ (VLDB) تقسیم می شوند.

پنج نسل ذخیره و بازیابی اطلاعات عبارتند از:

۱- نسل فایل‌های ساده ترتیبی

۲- نسل AM

۳- نسل DMS

۴- نسل DBMS

۵- نسل KBS

۱- نسل فایل‌های ساده ترتیبی

در این نسل از نوار به عنوان رسانه خارجی استفاده می شد. ساختار فایل‌ها ترتیبی بود و ساختار فیزیکی و منطقی فایل یکسان بود. تنها روش پردازش فایل‌ها، یکجا بود و نرم افزار واسطی برای مدیریت پردازش فایل‌ها وجود نداشت امکان اشتراکی کردن داده ها وجود نداشت و تکرار در ذخیره سازی داده ها در بالاترین حد بود. هر نوع تغییر در ساختار داده ها یا رسانه ذخیره سازی موجب بروز تغییر در برنامه می شد و نسخه های متعددی از یک فایل نگهداری می شد.

۲- نسل AM

در این نسل یعنی شیوه های دستیابی (Access Methods)، نرم افزاری به نام شیوه های دستیابی ایجاد شد و برنامه کاربر را از پرداختن به جنبه های فیزیکی محیط ذخیره سازی مستقل کرد. رسانه این نسل دیسک بود و امکان دسترسی ترتیبی و مستقیم به رکوردها (نه فیلدها) وجود داشت و تا حدودی ساختار فیزیکی و منطقی فایل‌ها از یکدیگر جدا شدند.

۳- نسل DMS

در این نسل یعنی نسل سیستم مدیریت داده ها ، نرم افزار کاملتری نسبت به نرم افزار دستیابی ایجاد شد و واسطه بین برنامه های کاربردی و فایل های محیط فیزیکی شد. در این نسل:

- ۱- از داده های اشتراکی در برنامه استفاده شد.
- ۲- آدرس دهی به داده ها در سطح فیلد ممکن شد.
- ۳- امکان بازیابی براساس چند کلید مهیاگشت.
- ۴- میزان افزونگی کاهش یافت.
- ۵- برنامه های کاربردی در قبال رشد فایلها مصون شدند.

۴- نسل DBMS

در نسل چهارم، برنامه های کاربردی از ویژگی های محیط فیزیکی ذخیره سازی مستقل شدند. کاهش افزونگی در ذخیره سازی داده ها، افزایش سرعت دستیابی به داده ها، بالا رفتن امنیت، امکان استفاده اشتراکی از داده ها و وجود یک محیط انتزاعی (Abstractive) از دیگر ویژگی های این نسل می باشد.

۵- نسل KBS

در این نسل یعنی نسل بانک معرفت (شناخت) به کمک مفاهیم هوش مصنوعی و سیستم های خبره، سیستمی طراحی شد که قادر به استنتاج منطقی از داده های ذخیره شده می باشد. در نسل پنجم سیستم بانک معرفت (KBS) ایجاد شد که مسئولیت ذخیره سازی شناخت، تامین جامعیت و امنیت بانک و تامین نیازهای کاربران را بر عهده دارد.

عناصر چهارگانه محیط پایگاه داده ها عبارتند از:

۱- نرم افزار (DBMS - نرم افزار شبکه - برنامه های کاربردی - رویه های ذخیره شده)

۲- سخت افزار (ذخیره سازی - ارتباطی - پردازشگر)

۳- کاربر (موردی (نامنظم) - همیشگی (منظم))

۴- داده

✓ به هر استفاده کننده از سیستم پایگاه داده ها، کاربر می گویند.

✓ کاربر منظم یا نامنظم خود بر دو نوع برنامه ساز و ناب برنامه ساز تقسیم می شوند. کاربر ناب برنامه ساز از طریق منو، هدایت می شود. کاربر برنامه ساز می تواند سیستمی یا کاربردی باشد.

✓ برنامه ساز کاربردی، برنامه های بهره برداری از پایگاه داده ها را می نویسد و برنامه ساز سیستم، برنامه های ایجاد و کنترل پایگاه داده ها را می نویسد.

✓ یک نوع کاربر نیز به نام کاربر پایانی (End-user) وجود دارد که به هر دو نوع کاربر برنامه ساز کاربردی و کاربر ناب برنامه ساز گفته می شود.

یک مدل داده ای شامل یک ساختار داده (DS) است. در واقع طراحی منطقی پایگاه داده ها به کمک مفاهیم اساسی یک مدل داده ای و در چار چوب ساختار داده ای آن مدل انجام می گیرد. ساختار داده ای امکانی است برای نمایش داده های موجودیت ها و انواع ارتباطات بین آنها.

عناصر تشکیل دهنده هر مدل

۱- ساختار داده ای

۲- امکانات عملیات در پایگاه داده ها

۳- امکانات کنترل جامعیت پایگاه داده ها

انواع ساختارهای داده ای

مدل های داده ای بر پایه رکورد بر سه نوع می باشند:

۱- رابطه ای (RDS)

۲- سلسله مراتبی (HDS)

۳- شبکه ای (NDS)

مدلهای داده ای بر پایه شیء عبارتند از: (ER , Semantic , Functional)

مفهوم ساختار داده ای بخشی از مفهوم مدل داده ای است.

ساختار رابطه ای (RDS) ، دارای ویژگی های زیر می باشد:

- ۱- مبنای تئوریک قوی دارد. (تامین کننده محیط انتزاعی به طور کامل)
- ۲- مسطح بودن محیط (مانند یک فایل ترتیبی ساده)
- ۳- دارای نمایش ساده از نظر کاربر
- ۴- دارای فقط یک عنصر ساختاری اساسی (جدول)
- ۵- امکان نمایش ارتباطات $1:1$, $1:N$, $N:M$
- ۶- ساده بودن منطق و دستور بازیابی
- ۷- دارای رویه پاسخگوی قرینه برای پرسشهای قرینه

ساختار سلسله مراتبی (HDS)، قدیمی ترین ساختار داده ای برای طراحی منطقی پایگاه داده ها می باشد که دو عنصر اساسی دارد: نوع رکورد و نوع پیوند پدر-فرزندی (PCL).

بین هر دو رکورد پشت سرهم در یک مسیر درخت، پیوند پدر فرزندی وجود دارد که ارتباط $1:N$ را نمایش می دهد. در این ساختار، هر رکورد فرزند، تنها یک رکورد پدر دارد، یعنی در یک نوع PCL شرکت دارد.

ویژگی های ساختار سلسله مراتبی عبارتند از:

- ۱- مبنای ریاضی ندارد.
- ۲- مناسب برای ارتباط $1:N$.
- ۳- سادگی نمایش ساختار رابطه ای را ندارد.
- ۴- نمایش ارتباط $N:M$ در آن مشکل است.

- ۵- جستجو حتماً باید از ریشه انجام شود.
 - ۶- نداشتن تقارن ساختار جدولی.
 - ۷- داشتن تعدادی محدودیت جامعیت.
 - ۸- مشکل بودن دستور بازیابی در آن نسبت به ساختار جدولی.
 - ۹- نمایش ارتباط با درجه بیشتر از دو.
- پایگاه داده سلسله مراتبی، مجموعه ای منظم از نمونه های یک یا چند نوع سلسله مراتب می باشد.
- در HDS، برای نمایش ارتباط چند به چند بین دو نوع موجودیت، می توان:
- ۱- دو نوع سلسله مراتب جدا طراحی کرد.
 - ۲- دو نوع سلسله مراتب به هم مرتب طراحی کرد.
 - ۳- یک نوع سلسله مراتب طراحی کرد.

ساختار شبکه ای (پلکس) توسط گروه DBGT پیشنهاد شد و اولین سیستم مدیریت پایگاه داده ای شبکه ای IDMS نام دارد. این سیستم گاهی به نام کوداسیل نامیده می شود. شبکه، نوعی گراف جهت دار است. در ساختار شبکه هر گره فرزند می تواند بیش از یک گره پدر داشته باشد، بنابراین گسترش یافته ساختار سلسله مراتبی است.

ساختار شبکه ای از دو عنصر تشکیل شده است:

۱- نوع رکورد

۲- نوع مجموعه (مجموعه کوداسیلی)

مجموعه کوداسیلی از سه جزء تشکیل شده است:

۱- نام مجموعه

۲- یک نوع رکورد مالک

۳- یک نوع رکورد عضو

درج مالک، بدون عضو، ممکن نمی باشد و درج یک عضو بدون وجود مالک ممکن نمی باشد.

با حذف مالک، اعضای آن نیز حذف می شوند.

ویژگی های ساختار شبکه ای

- ۱- تقارن دارد.
- ۲- مبنای ریاضی ندارد.
- ۳- سادگی ساختار رابطه ای را ندارد.
- ۴- عدم آنومالی در عملیات ذخیره سازی.
- ۵- فقط مخصوص نمایش ارتباطات $1:N$ نیست و هر نوع ارتباطی را می تواند نمایش دهد.
- ۶- امکان ناسازگاری داده ها در آن کمتر از سلسله مراتبی است.
- ۷- قواعد جامعیت ذاتی دارد.
- ۸- فزونکاری (به علت ایجاد و اصلاح اشاره گرها) دارد.
- ۹- دستور بازیابی پیچیده تری دارد.
- ۱۰- رویه جستجوی رکورد در آن نسبت به ساختارهای دیگر پیچیده تر است.
- ۱۱- اصل وحدت عملگر در یک عمل واحد مانند درج رعایت نمی شود.

در NDS، نوع رکورد مالک یک نوع مجموعه، می تواند عضو یا مالک در نوع مجموعه دیگر باشد.

در NDS اگر n نوع موجودیت و اشیانا دارای صفات چند مقداری، در یک نوع ارتباط شرکت داشته باشند، حداقل n مجموعه کوداسیلی برای طراحی پایگاه داده شبکه ای (NDB) لازم است.

در نمونه مجموعه مدل شبکه ای بر خلاف مجموعه ریاضی:

- ۱- حداقل یک عنصر وجود دارد.
- ۲- یک عنصر وجود دارد که از نظر نوع با عناصر دیگر فرق دارد.
- ۳- بین بعضی از عناصر نوعی نظم وجود دارد.

- ✓ اگر n نوع موجودیت در یک نوع ارتباط شرکت داشته باشند و این نوع ارتباط a صفت داشته باشد، رکورد پیوند دهنده در NDS، دارای $n+a$ فیلد در سطح پیاده سازی است.
- ✓ در NDS، صفات نوع ارتباط R با چندی $1:N$ بین دو نوع موجودیت E و F با فیلدهایی در نوع رکورد **عضو** نمایش داده می شوند.
- ✓ فقط در RDS، مفهوم نظم مطرح نیست.
- ✓ ساختار سلسله مراتبی تقارن ندارد، اما ساختار شبکه ای دارای تقارن است.
- ✓ تعداد عناصر ساختاری اساسی در RDS برابر 1 و در HDS و NDS برابر 2 می باشد.
- ✓ فقط در RDS، جستجو در پایگاه داده ها به صورت اتوماتیک (خودکار) است.
- ✓ با توجه به سه نوع ساختار رابطه ای، سلسله مراتبی و شبکه ای، سه دسته DBMS به ترتیب زمان پیدایش آنها به نام های HDBMS، NDBMS و RDBMS وجود دارد. بعد از آنها OODBMS است.

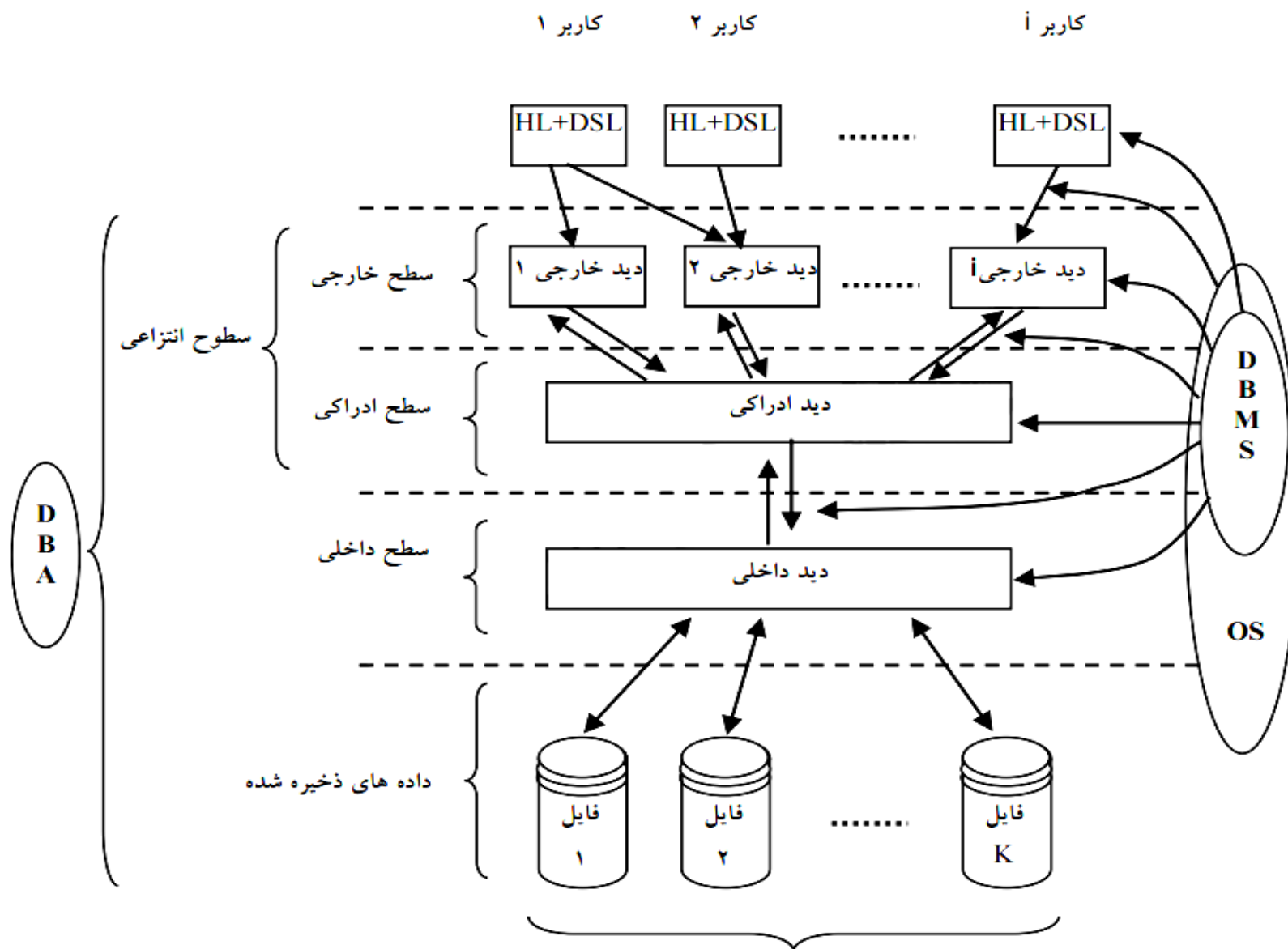
معماری استاندارد پایگاه داده ها که توسط ANSI پیشنهاد شد، یک معماری سه سطحی می باشد. این سطوح عبارتند از:

۱- سطح خارجی (External Level)

۲- سطح ادراکی یا مفهومی (Conceptual Level)

۳- سطح داخلی (Internal Level)

تذکر: به دو سطح خارجی و ادراکی، سطوح انتزاعی می گویند.



۱- دید ادراکی (Conceptual View)

دید، پنجره ای است که از آن کاربر می تواند محدوده پایگاه خود را ببیند و خارج از این محدوده، چیزی نمی بیند. دید ادراکی، دید طراح نسبت به داده های ذخیره شده در پایگاه است که در برگیرنده تمام نیازهای کاربران در محیط عملیاتی می باشد. دید ادراکی چون در یک محیط انتزاعی مطرح است، بر یک ساختار داده ای مشخص بنا شده است.

شمای ادراکی:

به توصیف و شرح دید ادراکی، شمای ادراکی می گویند. در واقع شمای ادراکی برنامه ای است شامل دستورات تعریف داده ها و کنترل داده ها (نه دستورات عملیات بر روی داده ها).

۲- دید خارجی (External View)

دید خارجی، دید کاربر نسبت به داده های ذخیره شده در پایگاه داده ها می باشد. این دید:

۱- یک دید جامع نمی باشد.

۲- روی دید ادراکی طراحی و تعریف می شود.

۳- در سطح انتزاعی مطرح است.

۴- مبتنی بر همان ساختار داده ای می باشد که دید ادراکی بر اساس آن طراحی می شود.

شمای خارجی:

به وصف یا تعریف دید خارجی، شمای خارجی می گویند. در واقع شمای خارجی برنامه ای است حاوی دستورات تعریف داده ها و گاه کنترل داده ها.

۳- دید داخلی (Internal View)

دید DBMS و طراح پایگاه داده ها نسبت به داده های ذخیره شده را دید داخلی می گویند.

شمای داخلی:

به تعریف دید داخلی، شمای داخلی می گویند. در واقع نوعی برنامه است که توسط خود DBMS تولید می شود. انواع رکوردها در شمای داخلی تعریف می شوند و شامل دستورهای لازم جهت ایجاد فایلها و کنترل آنها می باشد.

۴- کاربر

هر استفاده کننده از پایگاه داده ها را کاربر می نامند. کاربر می تواند موردی یا همیشگی باشد.

۵- زبان میزبان (HL : Host Language)

زبان میزبان می تواند یکی از زبانهای برنامه سازی چون پاسکال ، C ، ایدا ، ... باشد.

۶- زبان داده ای فرعی (DSL : Data Sub Language)

دستورهای این زبان به سه قسمت تقسیم می شود:

۱- دستورات تعریف داده ها (DDL)

۲- دستورات کنترل داده ها (DCL)

۳- دستورات عملیات روی داده ها (DML)