

معماری پایگاه داده

معماری سه سطحی ANSI/SPARC استاندارد برای طراحی سیستم های پایگاه داده است. کاربران مختلف در سطوح مختلف توسط یک زبان با سیستم تعامل می کنند.

ANSI/SPARC

سیستم های مدیریت پایگاه داده دارای معماری های یکسانی نیستند. معماری سه سطحی ANSI/SPARC یکی از استانداردهایی است که امروزه اساس اکثر سیستم های مدیریت پایگاه داده را شکل می دهد. این استاندارد توسط گروه مطالعاتی ANSI/SPARC اولین بار در سال ۱۹۷۵ برای طراحی سیستم های مدیریت پایگاه داده پیشنهاد شد.

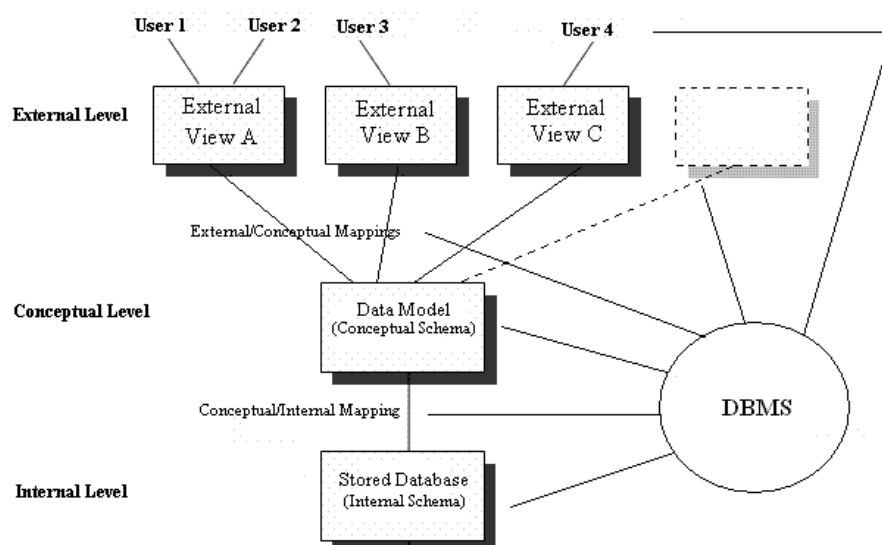
ANSI/SPARC مخفف American National Standards Institute, Standards Planning And Requirements Committee است.

معماری ANSI/SPARC سه سطح مجزا را برای توصیف داده در یک پایگاه داده تعیین می کند:

- سطح خارجی (external level)
- سطح ادراکی (conceptual level)
- سطح داخلی (internal level)

هدف معماری سه سطحی این است که امکاناتی را فراهم کند تا کاربران بتوانند با دیدگاه های شخصی خود به داده مورد نیاز دسترسی پیدا کنند. یعنی هر کاربری بتواند به داده مشترک دسترسی پیدا کند اما دید خاص خود را داشته باشد.

از طرف دیگر فاصله بین سطح داخلی از سطح خارجی دلالت بر این دارد که کاربر نیازی به دانستن جزئیات فیزیکی داده ذخیره شده در پایگاه داده ندارد. این تفکیک سطح اجازه تغییر ساختار ذخیره سازی پایگاه داده را بدون تاثیر روی دیدهای کاربران می دهد. لازمه این استقلال سطوح از همدیگر است به نحوی که تغییرات روی یک سطح روی بقیه تاثیر نگذارد.



سطح خارجی

سطح خارجی دید کاربر از داده های ذخیره شده در پایگاه داده است. منظور از دید کاربر (user view) قسمتی از پایگاه داده است که کاربر با آن سروکار دارد. یعنی مجموعه ای از صفات خاصه موجودیت هائی است که در اختیار کاربر قرار داده می شود. هر کاربر دیدگاه های خاص خود را از پایگاه داده می تواند داشته باشد.

دید هر کاربر باید تعریف شود. به تعریف و شرح دید کاربر شمای خارجی (external schema) می گویند. برای تعریف شمای خارجی از یک مدل داده استفاده می شود که معمولاً همان است که در سطح ادراکی بکار رفته است.

سطح ادراکی

سطح ادراکی کل داده های که در پایگاه داده ذخیره می شوند و ارتباط مابین آنها را شرح می دهد. یعنی داده هائی درباره انواع موجودیت ها و ارتباط آنها در محیط عملیاتی را توصیف می کند.

سطح ادراکی دید طراح پایگاه داده از داده های ذخیره شده در پایگاه داده است. داده های دنیای واقعی آنطور که واقعا هستند توسط طراح پایگاه داده مدل می شوند.

برای تعریف سطح ادراکی از یک ساختار یا مدل داده استفاده می شود که شمای ادراکی (conceptual schema) نامیده می شود. شمای ادراکی کلیه داده ها و ارتباط بین آنها را توصیف می کند. علاوه بر این رویه های شناسائی و قیدهای جامعیت را نیز دربر می گیرد.

برای کسب استقلال داده، شمای ادراکی تنها درگیر معنی داده است و جنبه های نمایش داده، سازماندهی فیزیکی و استراتژی های دستیابی ندیده گرفته می شود.

شمای خارجی از شمای ادراکی مشتق می شود و اگر مدل داده در هر دو سطح یکسان نباشد سیستم پایگاه داده را دوساختاری می نامند.

سطح داخلی

سطح داخلی درگیر چگونگی نمایش فیزیکی پایگاه داده روی سیستم کامپیوتری است و شرح می دهد چگونه داده واقعا در پایگاه داده و سخت افزار ذخیره می شود.

سطح داخلی دید طراح پایگاه داده از محیط فیزیکی ذخیره سازی و درواقع فایل های محیط فیزیکی است که توسط شمای داخلی (internal schema) توصیف می شود. شمای داخلی نحوه نمایش فیزیکی داده هائی را که در شمای ادراکی شرح داده شده را مشخص می کند. انواع مختلف رکوردها، فیلدهای داده، فایل ها، نحوه نمایش رکوردها در فایل، استراتژی دستیابی، شاخص ها و چگونگی ترتیب رکوردها در فایل توسط شمای داخلی تشریح می شوند.

جزئیات تبدیل به منبع ذخیره سازی در معماری سه سطحی بیان نمی شود و از این سطح به پایین در اختیار DBMS نیست و به عهده سیستم عامل و درایورهای دستگاه ذخیره سازی است.

مثال:

در یک سازمان موجودیت کارمند را در نظر بگیرید. دو کاربر User1 و User2 با دو دید مختلف با این موجودیت کار می کنند. تعریف دیدهای خارجی، ادراکی و داخلی می توانند چنین باشند:

خارجی.

User View 1(PL/1):
DCL 1 EMPP,
 2 Emp# Char(6),
 2 Sal Fixed Bin (31);
User View 2(COBOL):
01 EMPC.
 02 Empano Pic x(6).
 02 Pic x(4).

User 1		User 2	
Empano	Deptno	Emp#	Sal

ادراکی.

Employee
Employee_Number Character(6)
Department_Number Character(4)
Salary Numeric (5)

Employee		
Employee_Number	Department_Number	Salary

داخلی.

Stored_Emp
Prefix Type=Byte(6), Offset = 0
Emp# Type = Byte (6), Offset = 6, Index = Empx
Dept# Type = Byte (4), Offset = 12
Pay Type = FullWord, Offset = 16

تبدیلات بین سطوح

در معماری سه سطحی روش هائی برای تبدیل سطوح به یکدیگر وجود دارد. دو سطح از تبدیل موجود است:

- تبدیل ادراکی/داخلی ارتباط دید داخلی و ادراکی را تعریف می کند. تعیین می کند رکوردها و فیلدهای سطح ادراکی چگونه در سطح داخلی نمایش داده شوند. تغییرات در سطح داخلی باعث عوض شدن تبدیل ادراکی/داخلی می شود. بنابراین شیمای ادراکی ممکن

است ثابت بماند تا استقلال فیزیکی حاصل شود.

- تبدیل خارجی/ادراکی ارتباط دید ادراکی و یک دید خاص خارجی را برقرار می کند. مثلاً یک صفت خاصه از دید کاربری ممکن است ترکیبی از چندین فیلد در سطح ادراکی باشد. تغییر در تعاریف ادراکی باعث تغییر نگاشت ادراکی/خارجی می شود. از این رو شمای خارجی ممکن است ثابت بماند تا استقلال داده منطقی بدست بیاید.

استقلال داده

سیستم های قدیمی وابسته به داده بودند به این معنی که روش سازماندهی داده در دستگاه جانبی و روش دسترسی به آن توسط برنامه و در منطق آن ساخته می شدند. در چنین سیستمی تغییر در ساختار ذخیره سازی یا استراتژی دستیابی بدون تاثیر روی برنامه غیرممکن است.

یکی از مزایای سیستم پایگاه داده استقلال داده است. استقلال داده به معنی مصونیت برنامه ها از تغییرات ساختار ذخیره سازی و استراتژی دستیابی است. ساختار ذخیره سازی بدون تاثیر روی دیدهای کاربران است می تواند تغییر کند. پایگاه داده و برنامه های کاربری هم می توانند مستقل از یکدیگر تغییر کنند.

دو نوع استقلال داده وجود دارد:

۱. استقلال فیزیکی داده
 - استقلال فیزیکی داده نشان می دهد تا چه اندازه شمای داخلی می تواند بدون تاثیر روی برنامه های کاربردی تغییر کند.
۲. استقلال منطقی داده
 - استقلال منطقی داده نشان می دهد تا چه اندازه شمای ادراکی می تواند بدون تاثیر روی برنامه کاربردی تغییر کند.

زبان میزبان و زبان فرعی داده

برنامه نویسان، برنامه های کاربردی را با استفاده از یک زبان های سطح بالا نظیر Visual basic، Java و Visial C پیاده سازی می کنند. زبان سطح بالائی که علاوه بر داشتن امکانات گوناگون دارای دستوراتی برای تعریف و کار با داده هستند زبان میزبان (host language) نامیده می شوند.

زیرمجموعه ای از زبان میزبان که مختص عملیات ذخیره و بازیابی اطلاعات از پایگاه داده است زبان فرعی داده (Data Sub Language) نام دارد.

هر DSL ترکیبی از سه زبان دیگر است:

۱. احکام تعریف داده (DDL)
۲. احکام کارکردن با داده (DML)
۳. احکام کنترل (DCL)

زبان تعریف داده

DDL مخفف Data Definition Language امکان تعریف یا توصیف اشیای پایگاه داده را می دهد. ساختار رکوردها، تعریف فیلدها، محل فایل ها و شیوه ذخیره سازی داده ها در بانک به وسیله احکام DDL انجام می پذیرد.

مثال: نوع رکورد زیر را در نظر بگیرید.

```
create table account (  
    account-number char(10),  
    balance integer)
```

زبان دستکاری داده

DML مخفف Data Manipulation Language عملیات پردازشی و دستکاری اشیای پایگاه داده مانند insert، select، update را پشتیبانی می کند.

DML به عنوان زبان پرس و جو هم شناخته می شود و اغلب دارای قابلیت انجام محاسبات ریاضی و آماری است که عملیات گزارش گیری از پایگاه داده را آسان تر می کند.

زبان کنترل داده

DCL مخفف Data Control Language امکان تعیین نوع استراتژی های دستیابی، تعریف شاخص ها و مرتب سازی داده های پایگاه داده را می دهد.

دو دسته زبان DSL وجود دارد:

- رویه ای (Procedural). کاربر داده ای که نیاز دارد و نحوه دریافت آن را تعیین می کند.
- غیررویه ای (nonprocedural) یا (Declarative). کاربر تعیین می کند چه داده ای مورد نیاز است ولی نحوه حصول آن را بیان نمی کند.

هر سیستم پایگاه داده DSL خاص خود را دارد به عبارت دیگر هر مدل داده زبان فرعی خاص دارد. یک DSL خاص که توسط اغلب سیستم های فعلی پشتیبانی می شود SQL است. SQL یک زبان غیر رویه ای است.

سطوح داخلی، ادراکی و خارجی هریک DSL خاص خود را دارند. شمای هر سطح توسط DSL مربوطه نوشته می شود.

XML مخفف Extensible Markup Language که توسط کنسرسیوم W3C معرفی شده است زبان نشانه گذاری مستندات است تا زبان پایگاه داده. اما توانایی آن در تعیین تگ های جدید و تولید ساختارهای تودرتو باعث شد روش مهمی برای تبادل داده بشود و اکنون XML اساس کلیه فرمت های تبادل داده نسل جدید شده است. ابزارهای گوناگونی برای تجزیه، مرور و پرس و جو داده/مستندات XML موجود است.

کاربران پایگاه داده

کاربران یک سیستم پایگاه داده توسط روش هائی که با سیستم تعامل می کنند از هم تفکیک می شوند.

تحلیل گران سیستم

تحلیل گران سیستم (system analysts) با گروه کاربران پایگاه داده به منظور درک نیازهای اطلاعاتی و پردازشی آنها ارتباط دارند. نیاز های اطلاعاتی و پردازشی هر گروه را مجتمع می کنند و مستندسازی می کنند.

طراحان پایگاه داده

طراحان پایگاه داده (database designers) ساختار مناسبی را برای نمایش اطلاعات مشخص شده توسط تحلیل گر سیستم به طریق نرمالسازی شده به منظور تضمین جامعیت و سازگاری داده انتخاب می کنند و با استفاده از DDL داده های پایگاه داده را تعریف می کنند.

پیاده سازان برنامه های کاربردی

برنامه نویسان برنامه های کاربردی (Application Developers) برای برآوردن نیازهای کاربران و کار با پایگاه داده برنامه هایی را آماده می کنند. تست، اشکالزدائی و مستندسازی برنامه و پایگاه داده از وظایف برنامه نویسان است. برنامه نویسان با سیستم توسط احکام DML ارتباط برقرار می کنند.

مدیر پایگاه داده

مدیر پایگاه داده (database administrator) یا بطور خلاصه DBA فردی است که مسئول کنترل عملیات کل سیستم پایگاه داده است. DBA کلیه فعالیت های سیستم پایگاه داده را هماهنگ می کنند. این فرد باید درک خوبی از منابع و نیازهای اطلاعاتی کل سازمان داشته باشد و برای حصول اطمینان از اینکه داده موردنیاز قابل دسترس کاربران قرار می گیرد با آنها در ارتباط باشد.

بعضی از وظایف DBA شامل:

- تعریف شیماها توسط DDL
- تعریف ساختار ذخیره سازی و متدهای دسترسی توسط DDL
- اصلاح شیما و سازماندهی فیزیکی
- اعطای مجوز دسترسی پایگاه داده به کاربران
- تعیین قیدهای جامعیت
- عامل ارتباطی کاربران
- نظارت اجرا و واکنش برای تغییر در صورت نیاز
- برقراری دیکشنری داده

کاربران نهائی

کاربران نهائی (End Users) شامل:

- کاربران پارامتری: که توسط برنامه های کاربردی نوشته شده با سیستم سروکار دارند. مانند تحویلدار بانک و کارکنان دفتری.
- کاربران ماهر: که نیازهای پیچیده تری دارند و با قابلیت های DBMS آشنائی کامل دارند. درخواست های خود از پایگاه داده را توسط یک زبان پرس و جو می سازند.
- کاربران نهائی اتفاقی: کسانی که دسترسی گاه و بیگاه به پایگاه داده دارند اما ممکن است هر بار نیازهای متفاوتی داشته باشند. از زبان های پرس و جو و مرورگرهای حرفه ای تر استفاده می کنند.

دیکشنری داده

دیکشنری داده ها (Data Catalog) یکی از امکاناتی است که در سیستم پایگاه داده در اختیار DBA قرار می گیرد. دیکشنری داده ها که به آن راهنمای سیستم نیز می گویند یک متا داده است یعنی اطلاعاتی درباره خود پایگاه داده و داده های ذخیره شده در آن را نگهداری می کند.

دیکشنری داده تعیین می کند چه داده ای موجود است و چه معنی دارد، داده چگونه ذخیره می شود و در کجا قرار دارد، مالک آن چه کسی است و چه کسانی اجازه دسترسی به داده را دارند، تاریخچه و آمار استفاده از داده را دربردارد.