

به نام خدا

مقدمه‌ای بر تکنولوژی پایگاه داده‌ها

گردآوری:

سیدمحمد حسینی

منابع درس:

۱- مفاهیم بنیادی پایگاه داده‌ها

تألیف : سید محمدتقی روحانی رانکوهی

انتشارات جلوه، آخرین نگارش

۲- بانک اطلاعات علمی کاربردی

تألیف : دکتر مصطفی حق جو

انتشارات دانشگاه علم و صنعت ایران، آخرین نگارش

3- Database System Concepts

Abraham Silberschatz, Henry. F. Korth, S.Sudharsan,
5rd Edition, Tata McGraw Hill, 2006.

4- Fundamentals of Database Systems

Ramez Elmasri, Shamkant B. Navathe, 3rd Edition,
Addison Wesley, 2000.

5- An Introduction to Database Systems

DATE C.J., 7th Edition,
Addison Wesley USA 2000.

فهرست مطالب:

- ۱- مقدمه، کلیات و تاریخچه
- ۲- ساختارهای داده‌ای و معماری پایگاه داده
- ۳- سیستم مدیریت پایگاه داده
- ۴- مدل‌سازی معنایی داده‌ها
- ۵- مفاهیم مدل رابطه‌ای
- ۶- قواعد جامعیت
- ۷- جبر رابطه‌ای
- ۸- وابستگی و نرمال سازی

فصل اول

۱-۱- کلیات

اصطلاح پایگاه داده (Database) یکی از رایجترین اصطلاحات در دانش و فن کامپیوتر می- باشد، بطوریکه سالانه بیش از یکصد و پنجاه هزار صفحه مطلب آکادمیک و تکنیکی در این زمینه منتشر می شود.

سیستم مدیریت پایگاه داده ها (DBMS) یکی از سیستم های ذخیره و بازیابی اطلاعات است، با توجه به معنای عام سیستم ذخیره و بازیابی اطلاعات: "یعنی هر سیستم که به کاربر برنامه ساز یا ناب برنامه ساز امکان دهد تا اطلاعات خود را ذخیره، بازیابی، و پردازش کند و احیاناً اطلاعات جدیدی را تولید نماید."

پایگاه داده ها : امروزه همه کسانی که به نحوی با کامپیوتر سرو کار دارند، این اصطلاح را می شناسند و هر یک در حد درک خود، آنرا به کار می برند و بسیاری نیز در این زمینه آگاهی و مهارت و تجربه دارند.

ذخیره و بازیابی اطلاعات : مجموعه ای از الگوریتم ها و تکنیک ها که در ذخیره سازی، بازیابی و پردازش اسناد، مدارک، متون، تصاویر و اصوات به کار می روند. این گونه داده ها ممکن است ساختمانند، نیم ساختمانند^۱ یا حتی غیر ساختمانند^۲ باشند.

سیستم های مدیریت پایگاه داده ها نیز سیر تحول و تکامل خاص خود را دارند. این سیستم چندین رده تکنولوژی دارد:

۱- سیستم فایلینگ^۳

۲- سیستم مدیریت داده ها^۴

۳- سیستم مدیریت پایگاه داده ها^۵

۴- سیستم مدیریت پایگاه شناخت^۶

۵- سیستم مدیریت پایگاه داده های شیء گرا^۷

۶- سیستم هوشمند^۸ مدیریت پایگاه داده ها

^۱ Structured

^۲ Semi Structured

^۳ Unstructured

^۴ Filing Systems

^۵ Data Management System(DMS)

^۶ Database Management System(DBMS)

^۷ Knowledge Base Management System(KBMS)

^۸ Object Oriented Database Management System(OODBMS)

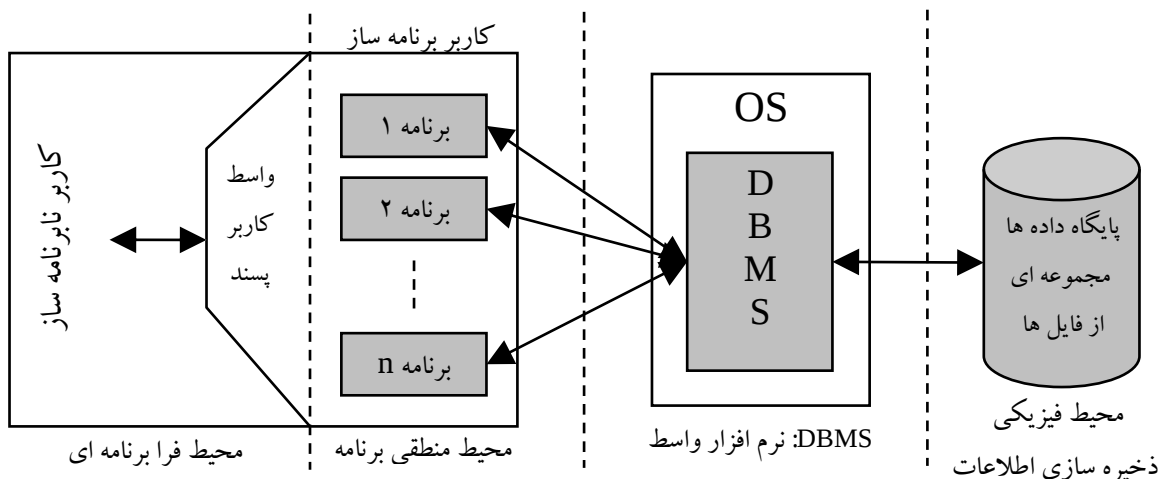
^۹ Intelegent System

۷- سیستم معنایی^۱ مدیریت پایگاه داده‌ها

۸- سیستم داده کاوی و کشف شناخت^۲

۹- سیستم مدیریت چند پایگاهی^۳

DBMS یک نرم‌افزار واسط ذخیره و بازیابی اطلاعات است.



۲-۱- تاریخچه

- تا پیش از دهه ۶۰، از سیستم‌های پرونده‌ای استفاده می‌شد.
- در اواسط دهه ۶۰، تکنولوژی پایگاه داده‌ها ایجاد گردید.
- در دهه ۷۰، توسعه یافته و به طور گسترده مورد استفاده قرار گرفت.
- اوایل دهه ۸۰، سیستم مدیریت پایگاه داده‌های رابطه‌ای به عنوان تکنولوژی برتر معرفی گردید.
- از اواسط دهه ۸۰ به بعد، سیستم‌های دیگری تولید و عرضه شدند، مثل سیستم مدیریت پایگاه شناخت، سیستم مدیریت پایگاه داده ای شیء‌گرا، و
- سیستم پرونده‌ای: در این روش اطلاعات راجع به یک موضوع را در فایل‌های مخصوص می‌گذاشتند و آنها را بر یک اساس خاص طبقه‌بندی می‌کردند. این روش دارای معایبی می‌باشد مثل : حجم بالا، سرعت پائین و ...
- سیستمی که بجای این روش آمده است ، سیستم بانک اطلاعاتی نام دارد.
- مزایای سیستم‌های بانک اطلاعاتی به سیستم‌های پرونده‌ای
- **فشرده‌گی:** نیاز به فایل های متنی حجیم نیست.
- **سرعت:** سرعت ماشین در محاسبات، ذخیره و بازیابی و ... نسبت به انسان بیشتر است.

¹ Semantic

² Data Mining and Knowledge Discovery System

³ Multi Database Management System(MDBMS)

- **بودجه کمتر:** سیستم های کامپیوتری یکبار ایجاد می شوند و بارها استفاده می شوند، بنابراین در نیروی انسانی، فضا و ... صرفه جویی می شود.
- **دسترسی:** یعنی بتوان در هر مکان و هر زمان به اطلاعات دسترسی داشت.
- **کنترل مرکزی:** سیستم بانک اطلاعاتی موجب می شود تا مؤسسه بر روی داده هایش کنترل مرکزی داشته باشد.
- مزایای روش های کنترل مرکزی:
- داده ها به صورت مشترک استفاده می شوند
- کاهش داده های اضافی
- کاهش ناسازگاری
- ایجاد جامعیت و یکپارچگی
- موازنه نیازمندی ها
- استانداردسازی امکانپذیر می شود

۳-۱- تعریف داده (Data)

- ۱- داده عبارتست از نمایش ذخیره شده اشیاء فیزیکی، چیزهای مجرد، بوده ها (واقعیات)، رویدادها یا موجودیت های دیگر قابل مشاهده که در تصمیم گیری بکار می روند.
- ۲- داده عبارتست از بوده یا هست معلوم که می توان بوده یا هست دیگری را از آن استنباط کرد.
- ۳- ANSI: نمایش بوده ها، پدیده ها، مفاهیم یا شناخته ها به طرز صوری و مناسب برای برقراری ارتباط، تفسیر یا پردازش توسط انسان یا هر امکان خودکار.
- ۴- ANSI: هر نمایشی اعم از کاراکتری یا کمیت های قیاسی که معنایی به آن قابل انتساب باشد.

۴-۱- تعریف اطلاع

- ۱- اطلاع به داده ای اطلاق می شود که توسط فرد یا سازمان برای تصمیم گیری بکار می رود.
- ۲- اطلاع داده پردازش شده است.
- ۳- اطلاع عبارتست از داده سازمان یافته که شناختی را منتقل می کند. به عبارت دیگر، اطلاع، داده ای است که سازمان یافته است تا چیزی را به ما بگوید.
- ۴- ANSI: اطلاع عبارتست از معنایی که انسان به داده منتسب می کند؛ از طریق قراردادهای شناخته شده ای که در نمایش داده بکار می روند.

۱-۵- پایگاه داده‌ها چیست؟

- ۱- پایگاه داده‌ها عبارتست از مجموعه‌ای از داده‌ها.
- ۲- پایگاه داده‌ها عبارتست از مجموعه‌ای از داده‌های منطقاً به هم مرتبط که برای پاسخگویی به نیازهای اطلاعاتی یک سازمان طراحی شده‌اند.
- ۳- پایگاه داده‌ها عبارتست از داده‌های نشان دهنده اشیائی از بخشی از جهان مورد استفاده در یک کاربرد مشخص.
- ۴- پایگاه داده‌ها عبارتست از مجموعه‌ای از داده‌های پایدار که در سیستم‌های کاربردی در یک سازمان مورد استفاده قرار می‌گیرند.
- ۵- پایگاه داده‌ها مجموعه‌ای است از داده‌ها، نوعاً نشان دهنده فعالیت‌های یک یا چند سازمان به هم مرتبط.

تعریفی جامع

۶- بانک اطلاعاتی مجموعه‌ای است از داده‌های ذخیره شده و پایدار، به صورت مجتمع^۱ و مبتنی بر یک ساختار، به هم مرتبط، تعریف شده به صورت صوری و با کمترین افزونگی^۲، تحت مدیریت یک سیستم کنترل متمرکز^۳، مورد استفاده یک یا چند کاربر از یک یا بیش از یک سیستم کاربردی به طور همزمان^۴ و اشتراکی^۵.

داده پایدار: داده‌ای است که وقتی توسط سیستم مدیریت بانک برای ورود به بانک پذیرفته شد، فقط در صورتی می‌تواند تغییر کند که درخواستی در این رابطه به سیستم ارسال شود؛ یا به عبارت دیگر، با اثرات جانبی ناشی از اجرای برنامه حذف نشود.

مجتمع^۱ و مبتنی بر یک ساختار: به معنی آن است که کل داده‌های عملیاتی محیط مورد نظر کاربران مختلف، در قالب یک ساختار مشخص بصورت یکجا ذخیره شده باشند. به عبارتی پراکندگی در ذخیره سازی داده‌های محیط وجود نداشته باشد.

تعریف شده بصورت صوری: به معنی آن است که داده‌ها به کمک احکام خاصی، در کادر تعریف فایل‌های مورد نیاز، تشریح و تعریف شوند و این کار زبان خاصی را لازم دارد.

¹ Integrated
² Redundancy
³ Centralized Control
⁴ Concurrent
⁵ Shared
⁶ Integrity

مثالی برای درک بهتر مفهوم پایگاه داده

محیط عملیاتی دانشگاه را در نظر بگیرید که دارای بخش های عملیاتی مختلف می باشد. فرض می شود که سه بخش امور آموزش، امور دانشجویی و امور مالی دانشگاه بخش هایی هستند که می خواهیم برای آنها یک سیستم ذخیره و بازیابی کامپیوتری ایجاد نماییم و نیز فرض می کنیم که تنها نوع موجودیت مورد نظر، موجودیت دانشجو باشد و بخش های فوق می خواهند اطلاعاتی را در مورد این نوع موجودیت داشته باشند. واضح است که در هر یک از بخش های فوق انواع دیگری از موجودیت ها وجود دارند که در این مثال مورد بحث قرار نمی گیرند.

برای ایجاد سیستم هایی که نیاز به نگهداری اطلاعاتی در رابطه با موضوعی خاص می باشند دو روش کلی وجود دارد:

۱- روش پرونده ای (سنتی / فایلینگ)

۲- روش پایگاهی

روش پرونده ای یا غیر بانکی (سیستم فایل پردازی)

در این روش هر یک از زیر محیط های عملیاتی مستقلاً مطالعه می شود و برای هر زیر مجموعه، یک سیستم خاص همان زیر محیط طراحی و تولید می شود، بگونه ای که فقط پاسخگوی همان زیر محیط است. با توجه به مثال مطرح شده هر قسمت از دانشگاه سیستم کاربردی خاص و جداگانه خود را خواهد داشت.

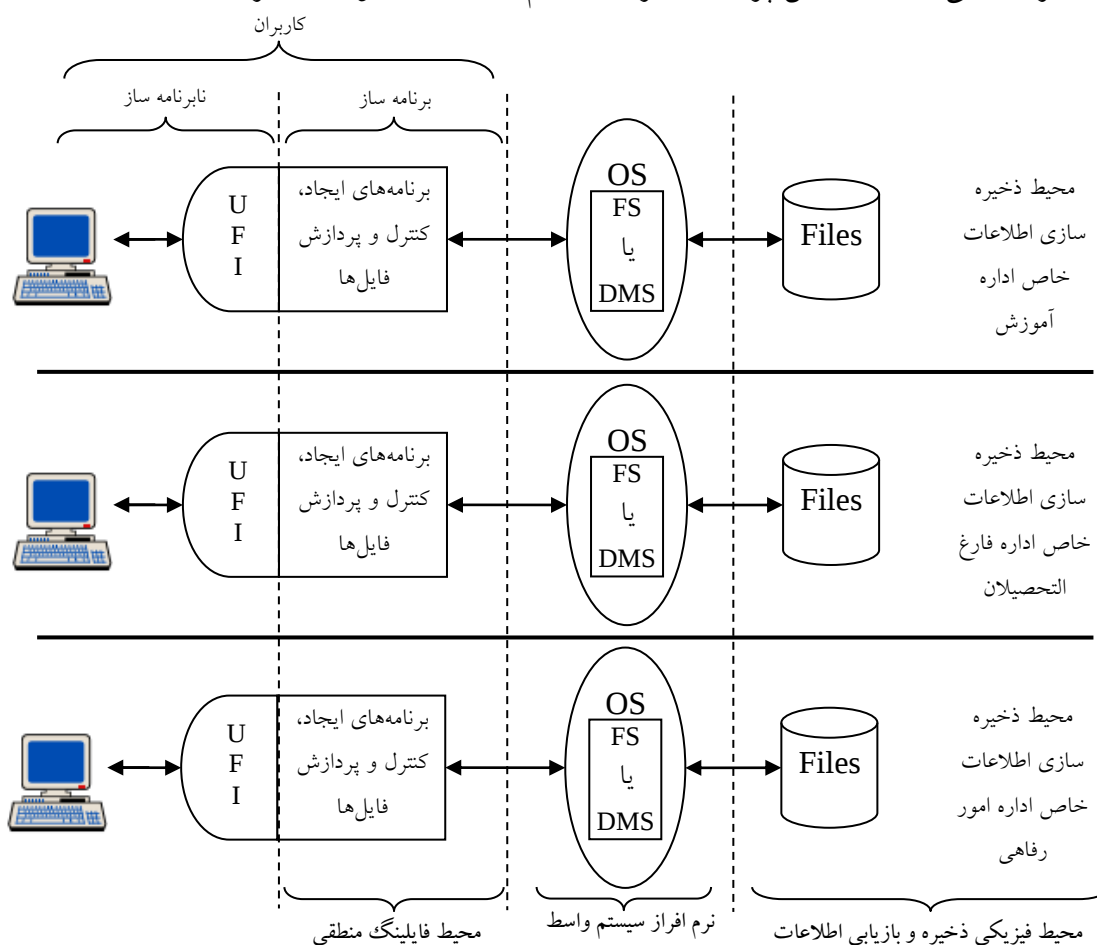
مشخصه های این روش عبارتند از:

- ۱- در روش فایل پردازی داده ها از هم مجزا می باشند
- ۲- محیط ذخیره سازی نامجموع است (تعدادی سیستم جداگانه و محیط ذخیره سازی جداگانه)
- ۳- برنامه های کاربردی وابسته به قالب فایل می باشند
- ۴- عدم سازگاری در داده ها و فایلها دیده می شود
- ۵- اشتراکی نبودن داده ها : داده های زیر محیط ۱ مورد استفاده کاربران زیر محیط ۲ نمی توانند قرار گیرند
- ۶- اطلاعات تکراری و افزونگی در داده ها وجود دارد

¹ File Processing

- ۷- عدم امکان استانداردهای واحد محیط عملیاتی بدلیل وجود سیستم های متعدد و پراکنده که
 گاهی اوقات توسط تیم های مختلف طراحی و پیاده سازی شده است و امکان اعمال یکسری از
 عملیات منسجم روی آن سیستم وجود ندارد
- ۸- مصرف غیر بهینه امکانات سخت افزاری و نرم افزاری و حجم زیاد برنامه سازی و
 استفاده غیر بهینه از مهارت و وقت تیمهای برنامه سازی

در زیر نمایی ساده از روش پرونده ای برای سیستم دانشگاه (به طور مختصر) آورده شده است:



قالب رکورد از دید آموزش

دانشکده، سال ورود، تاریخ تولد، نام خانوادگی، نام، شماره دانشجوی

قالب رکورد از دید امور دانشجویی

سال ورود، تاریخ تولد، نام خانوادگی، نام، شماره دانشجوی، تاریخ فارغ التحصیلی

روش پایگاهی یا بانکی^۱

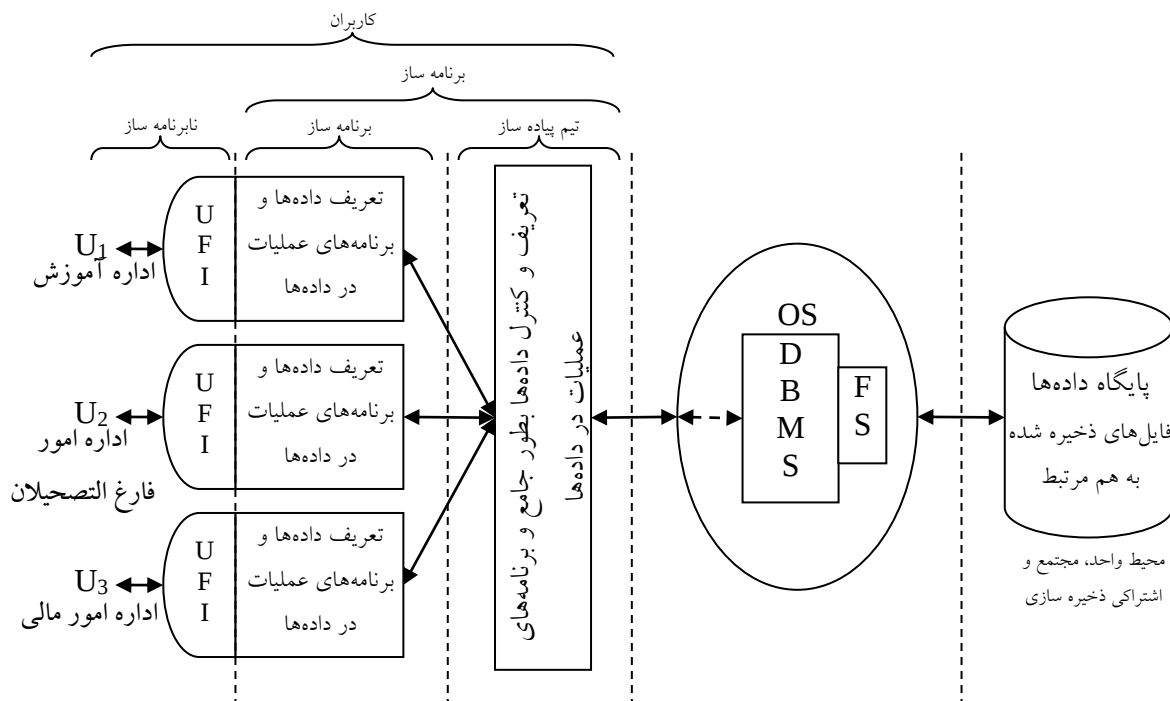
در این روش یک تیم واحد طراحی و پیاده سازی به سرپرستی متخصصی به نام DBA مجموعه نیازهای اطلاعاتی کل محیط عملیاتی مورد نظر مدیریت کل سازمان را بررسی می کند و با توجه به نیازهای اطلاعاتی تمام کاربران محیط و ضمن استفاده از یک نرم افزار خاص به نام DBMS محیط واحد و مجتمع ذخیره سازی اطلاعات ایجاد می شود. با توجه به مثال مطرح شده، رکورد نوع دانشجو فقط یکبار در فایل ذخیره می شود و کاربران مختلف هر یک طبق نیاز اطلاعاتی خود از آن بطور مشترک استفاده می نمایند. در رکورد نوع دانشجو، تمام صفات خاصه مورد نیاز کاربران مختلف وجود دارند و صفات خاصه مشترک، تنها یکبار در رکورد منظور می شوند. در این روش هر کاربری، دید خاص خود را نسبت به داده های ذخیره شده در بانک دارد. دید کاربران مختلف از یکدیگر متفاوت و حتی گاه با هم متضاد است.

مشخصه های این روش

- ۱- داده های مجتمع: کل داده ها بصورت یک بانک مجتمع دیده می شوند و از طریق DBMS با آنها ارتباط برقرار می شود.
- ۲- عدم وابستگی برنامه های کاربردی به داده ها و فایلها: زیرا DBMS خود به مسائل فایلینگ می پردازد و کاربران در محیط انتزاعی هستند.
- ۳- تعدد شیوه های دستیابی به داده ها
- ۴- عدم وجود ناسازگاری در داده ها
- ۵- اشتراکی بودن داده ها
- ۶- امکان ترمیم داده ها
- ۷- کاهش افزونگی
- ۸- کاهش زمان تولید سیستم ها
- ۹- امکان اعمال ضوابط دقیق ایمنی

^۱ Database Approach

در زیر نمایشی ساده از روش پایگاهی برای سیستم دانشگاه (به طور مختصر) آورده شده است:



۶-۱- عناصر محیط پایگاه داده

محیط پایگاه داده‌ها از چهار عنصر تشکیل شده است:

۱- سخت‌افزار

۱-۱- سخت‌افزار ذخیره سازی داده‌ها: منظور همان رسانه‌های ذخیره سازی است که

معمولاً برای ذخیره سازی داده ها از دیسکهای سریع با ظرفیت بالا

استفاده می شود.

۲-۱- سخت‌افزار پردازشگر: منظور همان کامپیوتر یا ماشین است. ماشینهای خاص برای

محیطهای بانک اطلاعاتی نیز طراحی و تولید شده اند

که به نام DBM ماشینهای بانک اطلاعاتی نیز خوانده می شوند. این ماشینها از

نظر معماری، حافظه اصلی،... و سایر اجزاء از ویژگیها و جنبه هایی برخوردارند که

شرح آن در این جزوه نمی گنجد.

۳-۱- سخت‌افزار ارتباطی: منظور مجموعه امکانات سخت افزاری است که برای برقراری

ارتباط بین کامپیوتر و دستگاههای جانبی و نیز بین دو

کامپیوتر یا بیشتر بکار گرفته می شوند، اعم از اینکه ارتباط نزدیک باشد و یا

ارتباط دور. سخت افزار ارتباطی خاص محیط های بانکی نیست و در هر محیط غیر بانکی نیز ممکن است مورد نیاز باشند.

۲- نرم افزار

۱-۲- سیستم مدیریت پایگاه داده ها (DBMS): که در محیط سیستم عامل اجرا می شود. بین بانک اطلاعاتی فیزیکی که داده ها بصورت فیزیکی در آن ذخیره می شوند و کاربران سیستم، لایه ای از نرم افزار موسوم به مدیر بانک اطلاعاتی قرار دارد. سیستم مدیریت بانک اطلاعاتی نرم افزاری است که به کاربران امکان می دهد که پایگاه از دید خود را تعریف کنند و به پایگاه خود دسترسی داشته باشند، با پایگاه خود کار کنند و روی آن کنترل داشته باشند.

۲-۲- برنامه های کاربردی: نرم افزاری است که کاربر باید برای تماس با سیستم بانک اطلاعاتی آماده کند.

۳-۲- رویه های ذخیره شده

۴-۲- نرم افزار شبکه: در صورتی که نیاز باشد پایگاه داده ها در محیط شبکه مورد استفاده قرار گیرد.

۳- کاربر

۱-۳- برنامه نویسان کاربردی: مسئول نوشتن برنامه های کاربردی بانک اطلاعاتی است که به زبان خاصی نوشته شده است می باشند.

۲-۳- کاربران نهایی: کسانی هستند که از نتایج بانک استفاده می کنند.

۳-۳- مدیر بانک: شخصی است که مسئول و هماهنگ کننده بانک است.

۴- داده

۱-۴- داده های ورودی: داده هایی که برای اولین بار وارد کامپیوتر می شوند.

۲-۴- داده های خروجی: داده هایی که بعد از پردازش به عنوان نتیجه کار گزارش می شوند.

۳-۴- داده های عملیاتی: داده هایی که با داده های ورودی و خروجی فرق می کنند و پس از یکسری عملیات به دست آیند. منظور داده هایی است که در مورد انواع موجودیتهای محیط عملیاتی و ارتباط بین آنها می باشند که به آنها داده های عملیاتی و یا داده های پایا گفته می شود. داده های ذخیره شدنی در پایگاه داده ها ابتدا باید در بالاترین سطح انتزاع مدلسازی معنایی شوند. مفاهیم داده ها در هر محیط به کمک موجودیت ها و ارتباطات نمایش داده می شوند.