

به نام خدا



اصول و مبانی فناوری اطلاعات

فصل سوم

پایگاه داده، انبار داده و داده کاوی

Chapter

3

Database, Data
Warehouse, and
Data Mining



بخش اول پایگاه داده و انبار داده

- هرچه شتاب تولید داده توسط سازمان‌ها افزایش می‌یابد، تکنولوژی ذخیره‌سازی، تحلیل و گزارش‌گیری داده‌ها نیز باید بهبود یابد و سریع‌تر و قدرتمندتر گردد.
- پایگاه‌های داده (databases)، انبارهای داده (data warehouse)، مارتهای داده (data marts) و هوش تجاری (business intelligence) این امکان را فراهم می‌آورد و به مدیران اجازه می‌دهد با قاطعیت، سرعت و اطمینان تصمیم‌گیری نمایند.
- علاوه بر داده‌های دیجیتال، حجم مستندات فیزیکی نیز هر سال افزایش می‌یابد. تکنولوژی مدیریت اسناد (document management) آرشیوی از اسناد دیجیتال و فیزیکی را به منظور برآوردن نیازهای سازمان‌ها تهیه می‌نماید.

تکنولوژی پایگاه داده

5

- Database: برای ذخیره سازی داده هایی که توسط برنامه های کاربردی، سنسورها، سیستم های پردازش تراکنش (TPS) تولید شده اند.
- Data warehouse: از منابع مختلفی مانند پایگاه های داده و سیلوهای داده، داده ها را جمع کرده و برای تحلیل های پیچیده تر، کشف دانش و استفاده در برنامه های تصمیم یار سازماندهی می نماید.
- Data mart: نوع کوچکی از انبارای داده است.
- Business intelligence: ابزارها و تکنیک هایی برای پردازش داده ها و تحلیل های آماری است و هدف آن کشف دانش و ارتباط بین داده ها و شناسایی ریسک و فرصت ها است.

- وقتی تراکنشی تجاری انجام میشود، مثلاً کالایی فروخته یا برگشت داده می‌شود، سفارشی پذیرفته یا ارسال میگردد، حقوقی پرداخته یا سپرده‌ای گذاشته می‌شود، تغییرات به‌صورت بلادرنگ در پایگاه داده اعمال می‌گردد.
- این تغییرات آنلاین شامل: اضافه نمودن به، بروز رسانی و حذف نمودن از پایگاه داده می‌باشد.
- سیستم مدیریت پایگاه داده (DBMS) تراکنش‌ها را در پایگاه داده ذخیره و پردازش می‌کند، و پرس و جو و گزارشگیری را اعمال می‌نماید.
- با توجه به عملکردی که DBMS ها دارند به آنها سیستمهای پردازش تراکنش آنلاین (online transaction processing (OLTP) systems) نیز گفته می‌شود.
- OLTP پایگاه داده‌ای است که کارهای پیچیده را به زیربخش‌های ساده‌تری می‌شکند. پایگاه داده‌های OLTP قادرند میلیون‌ها تراکنش را در هر ثانیه انجام دهند.



Figure 3.3 Data generated by transactions via computers, handhelds, and other input devices are stored in databases. DBMS support data queries and reports by extracting data from the database and delivering it to a screen or other output device.

برخی نرم افزارهای پایگاه داده

Oracle ☐

■ حدود ۴۸,۸٪ از سهم بازار را در اختیار دارد.

DB2 ☐

■ از رقبای جدی Oracle است.

SQL Server ☐

■ سادگی و سهولت اجرا و سازگاری کامل آن ویندوز از مشخصات این DBMS است.

PostgreSQL ☐

■ پیشرفته‌ترین پایگاه داده متن باز است که اغلب توسط بازی‌های آنلاین، Yahoo!، Skype و MySpace استفاده می‌شود.

MYSQL ☐

■ در بسیاری از وبسایتهای تجاری و کاربردهای داخل سازمانی مورد استفاده قرار می‌گیرد.

ORACLE®
DATABASE



انواع پایگاه داده :

پایگاه داده متمرکز (Centralized Database)

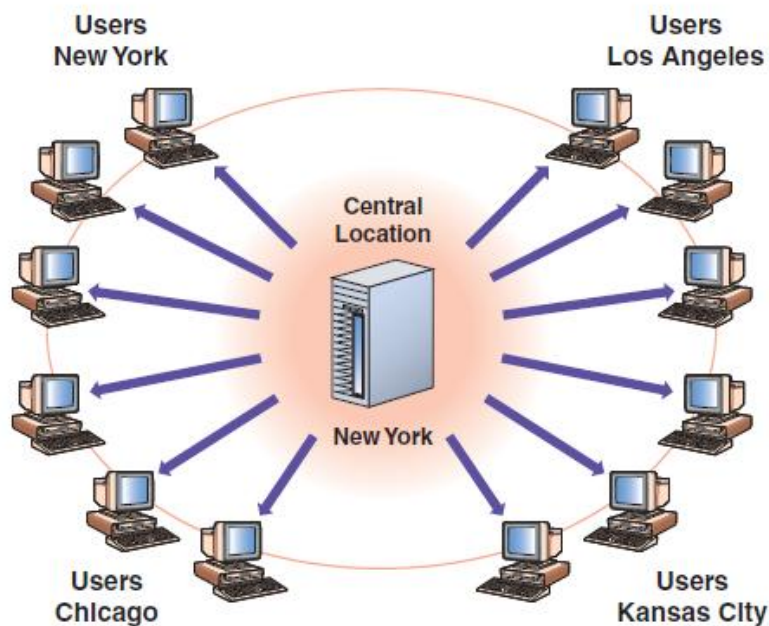
□ همه فایل‌ها در یک مکان فیزیکی مرکزی ذخیره می‌شوند.

□ فواید : امنیت و سازگاری داده‌ها

□ مشکلات :

□ کاهش سرعت انتقال اطلاعات هنگامی که کاربران از نظر جغرافیایی پراکنده باشند.

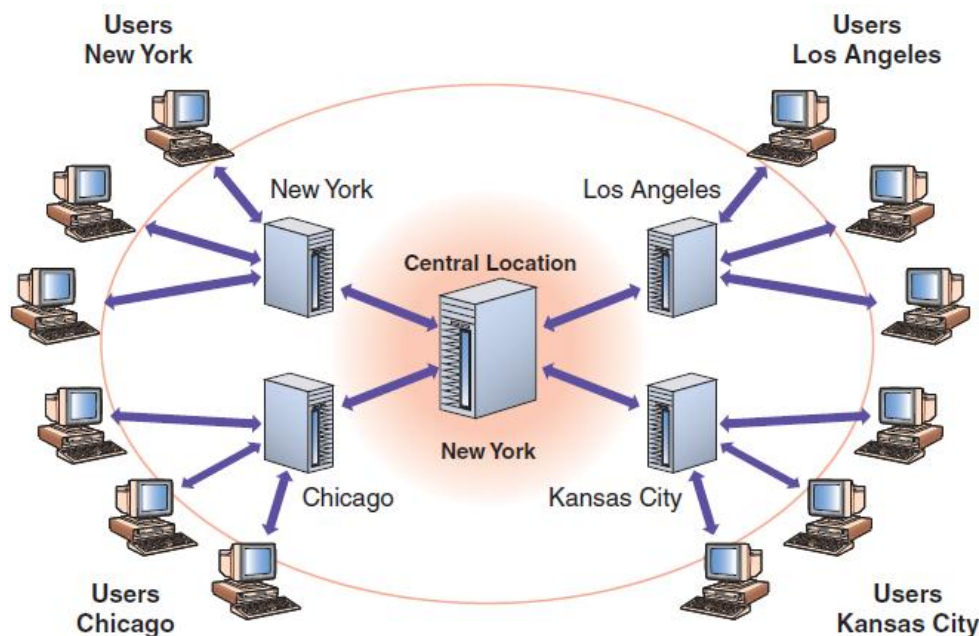
□ بروز مشکل برای همه کاربران، اگر کامپیوتر پایگاه داده متمرکز با مشکلی روبه‌رو شود



انواع پایگاه داده :

پایگاه داده توزیع شده (Distributed Database)

- پایگاه داده توزیع شده، دارای نسخه کاملی از پایگاه داده در چندین محل (تکثیر شده/Replicated) و یا داشتن بخشی از پایگاه در هر محل (افراز شده/Partitioned) می باشد
- امکان دسترسی برنامه های کاربردی به پایگاه داده های محلی و دور وجود دارد.



- سیستم به صورت کلاینت و سرور می باشد که کلاینت ها سیستم مشتری ها و پایگاه داده ها در سیستم های سرور قرار دارد که می تواند روی مرکز داده یک شرکت، یا در فضای ابری باشد.
- مزایا و مشکلات بر خلاف مدل قبل است

سیستم مدیریت پایگاه داده (DBMS)

10

جدول ۲-۳ مزایا و قابلیت‌های سیستم مدیریت پایگاه داده

ماندگاری. داده‌ها به‌طور پیوسته روی هارددرایو خارجی یا رسانه ذخیره‌سازی سریع و قابل اعتماد دیگری ذخیره می‌شوند تا مشخصاً حذف شوند یا تغییر داده شوند.

پرس‌وجو^۳. فرآیند پرس‌وجو داده‌ها را از منظرهای مختلف درخواست می‌کند؛ مثلاً «چند کامیون در تگزاس سبز هستند؟» **هم‌زمانی.** ممکن است افراد بسیاری سعی در خواندن یا تغییر داده‌های یکسانی را در یک زمان خاص داشته باشند. این داده‌ها ممکن است بدون وجود قوانین به‌اشتراک‌گذاری تغییرات، دارای ناسازگاری یا گمراه‌کننده شوند. به‌عنوان مثال اگر شما صفت رنگ ماشین ۷ را به «آبی» تغییر داده و در همان زمان فرد دیگری آن را به «قرمز» تغییر دهد، نتایج غیرقابل پیش‌بینی می‌شوند. سیستم‌های مدیریت پایگاه داده ابزارها و تکنیک‌های گوناگونی را برای مواجهه با چنین شرایطی فراهم می‌آورند.

پشتیبان‌گیری و کپی‌برداری. در صورت خرابی تجهیزات باید نسخه‌های پشتیبان را تهیه کرد.

اعمال قانون. قوانین برای پاک و قابل اعتماد نگه داشتن داده‌ها اعمال می‌شوند. به‌عنوان مثال یک قانون می‌تواند بیان کند که هر اتومبیل تنها اجازه دارد دارای یک موتور باشد (که با شماره موتور شناسایی می‌شود). اگر فردی سعی در مرتبط کردن موتور دیگری را هم داشته باشد، سیستم مدیریت پایگاه داده آن را متوقف کرده و پیام خطایی را نمایش می‌دهد. با این حال با وجود اتومبیل‌های هیبریدی گازی-برقی چنین قوانینی باید منعطف و اصلاح شوند. قوانین می‌توانند بدون طراحی مجدد قابل‌ملاحظه‌ای بنا بر نیاز حذف یا اضافه شوند.

امنیت. محدودیت روی افرادی که می‌توانند صفات را مشاهده کرده یا آن‌ها را تغییر دهند باید اعمال گردد.

محاسبه. به‌جای آن که برنامه‌های کاربردی هر کامپیوتر محاسبات را انجام دهند، سیستم‌های مدیریت پایگاه داده این وظیفه را بر عهده می‌گیرند.

ثبت تغییر و دستیابی. سیستم مدیریت پایگاه داده رکوردی را ایجاد کرده و اطلاعاتی نظیر چه کسی به چه صفاتی دسترسی داشته و چه چیزی را و در چه زمانی تغییر دهد، ممیزی می‌کند.

بهینه‌سازی خودکار. اگر الگوها یا درخواست‌های استفاده مکرری وجود داشته باشند، بسیاری از سیستم‌های مدیریت پایگاه داده می‌توانند به‌گونه‌ای تنظیم شوند که زمان پاسخ‌گویی را بهبود دهند.

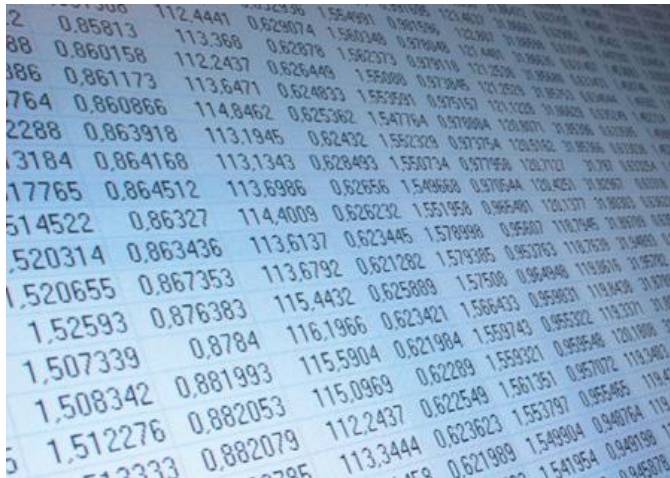
□ DBMS: برنامه‌ای که دستیابی به پایگاه‌های داده را فراهم می‌کند.

سیستم مدیریت پایگاه داده (DBMS)

وظایف :

- فیلتر کردن و مشخص کردن داده ها (Data Filtering and Profiling)
 - بررسی داده‌ها برای خطا، ناسازگاری، تکرار و اطلاعات ناکامل
- جامعیت داده ها (Data Integrity)
 - تصحیح، استانداردسازی و تایید جامعیت داده‌ها
- هماهنگ سازی داده ها (Synchronization)
 - تجمیع و جفت‌سازی داده‌های منابع مختلف
- غنی سازی داده‌ها (Enrichment)
 - استفاده از منابع داخلی و خارجی برای غنی‌سازی داده‌ها
- نگه داری داده ها (Maintenance)
 - بررسی و کنترل مداوم جامعیت داده‌ها

- بهترین استفاده از پایگاه داده‌ها پاسخ به پرس‌وجو و کوئری‌هایی است که کاربر درخواست نموده است.
- داده‌ها از طریق منابعی مانند پایانه‌های فروش (POS)، اسکنرها، فروشندگان آنلاین و غیره در قالبی ساختاریافته (structured) با توجه به نوع DBMS در پایگاه داده ذخیره می‌شوند.
- پایگاه داده‌های سنتی برای مدیریت داده‌های نامرتب مانند tweetها و یادداشتها مناسب نمی‌باشند.
- پایگاه داده‌ها برای داده‌کاوی، پردازش‌های تحلیلی برخط (online analytics processing- OLAP) پیچیده و سیستم‌های تصمیم‌یار مناسب نیست.
- این محدودیت‌های پایگاه داده‌ها منجر به معرفی انبارها و مارت‌های داده شده است که مناسب برای OLAP، داده‌کاوی، BI و تصمیم‌یار است.



2	0.85813	112.4441	0.629074	1.554991	0.901956	123.4567	0.123456	0.123456	0.123456
88	0.860158	113.366	0.62878	1.562373	0.979116	121.4801	0.123456	0.123456	0.123456
886	0.861173	112.2437	0.626448	1.550088	0.979116	121.4528	0.123456	0.123456	0.123456
764	0.860866	113.6471	0.624833	1.553591	0.975157	121.2523	0.123456	0.123456	0.123456
2288	0.863918	114.8462	0.625362	1.547764	0.979084	120.1071	0.123456	0.123456	0.123456
3184	0.864168	113.1343	0.628493	1.550734	0.977958	120.2702	0.123456	0.123456	0.123456
17765	0.864512	113.6986	0.62656	1.549668	0.970544	120.4251	0.123456	0.123456	0.123456
514522	0.86327	114.4009	0.626232	1.551958	0.965401	120.1377	0.123456	0.123456	0.123456
520314	0.863436	113.6137	0.623445	1.578998	0.956807	118.7945	0.123456	0.123456	0.123456
1520655	0.867353	113.6792	0.621282	1.579385	0.953763	118.7638	0.123456	0.123456	0.123456
1520655	0.867353	115.4432	0.625889	1.575308	0.964948	118.8814	0.123456	0.123456	0.123456
152593	0.876383	116.1966	0.623421	1.568433	0.959374	118.4408	0.123456	0.123456	0.123456
1507339	0.8784	115.5904	0.621984	1.559321	0.957072	118.3403	0.123456	0.123456	0.123456
1508342	0.881993	115.0969	0.62289	1.561351	0.957072	118.3403	0.123456	0.123456	0.123456
1512276	0.882053	112.2437	0.622549	1.561351	0.957072	118.3403	0.123456	0.123456	0.123456
1512333	0.882079	113.3444	0.623623	1.553797	0.957072	118.3403	0.123456	0.123456	0.123456

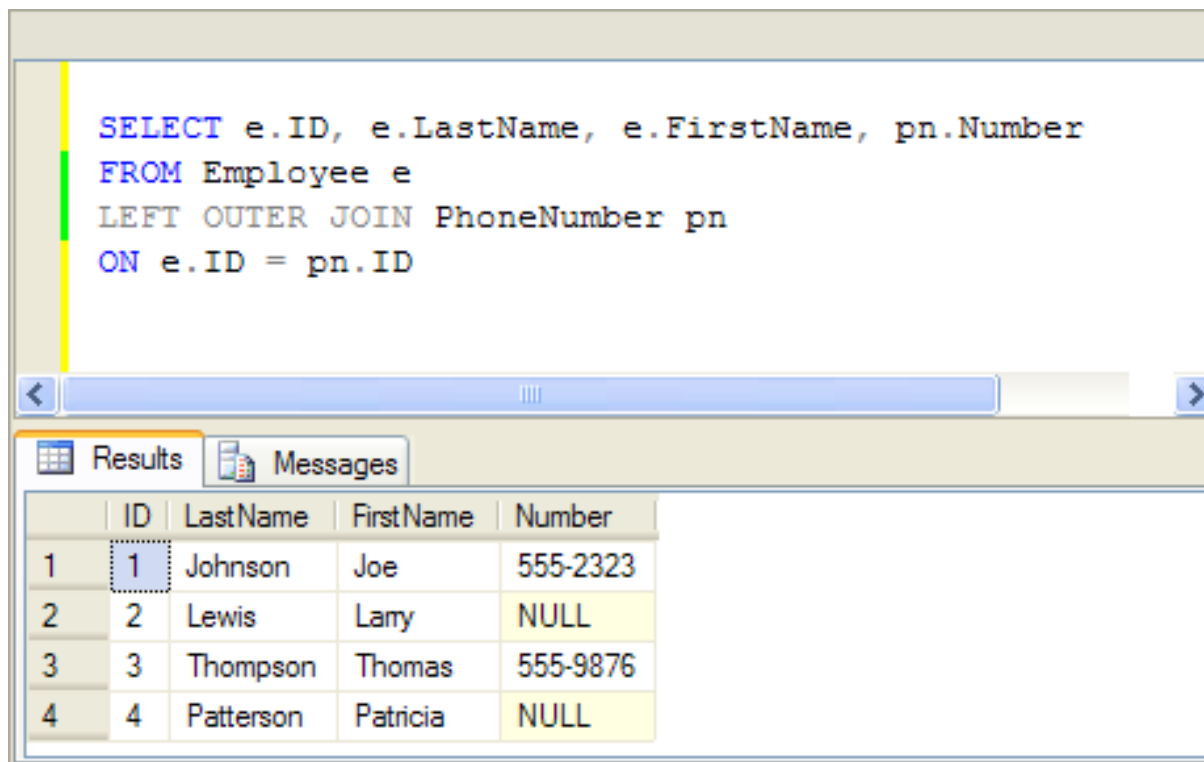
Figure 3.6 Illustration of structured data format. Numeric and alpha-numeric data are arranged into rows and predefined columns similar to an Excel spreadsheet.



Figure 3.7 Database queries are processed in real time and results are transmitted via wired or wireless networks to computer screens or handhelds.

پرس و جو در پایگاه داده

□ Queries



```
SELECT e.ID, e.LastName, e.FirstName, pn.Number
FROM Employee e
LEFT OUTER JOIN PhoneNumber pn
ON e.ID = pn.ID
```

The screenshot shows a SQL query execution window. The query is a LEFT OUTER JOIN between the Employee table (e) and the PhoneNumber table (pn) on the condition e.ID = pn.ID. The results are displayed in a table with 5 columns: ID, LastName, FirstName, and Number. The first row (ID 1) is highlighted with a blue selection box around the ID cell. The second and fourth rows (IDs 2 and 4) have a yellow background, indicating NULL values in the Number column.

	ID	LastName	FirstName	Number
1	1	Johnson	Joe	555-2323
2	2	Lewis	Lary	NULL
3	3	Thompson	Thomas	555-9876
4	4	Patterson	Patricia	NULL

انبار داده (Data warehouse)

- انبار داده نوع خاصی از پایگاه داده است که داده ها از پایگاه داده های مختلف به طور خلاصه شده و فقط خواندنی در آن ذخیره می شوند. این کار سرعت گزارش گیری را بسیار بالا می برد.
- در واقع انبار داده، یک بانک اطلاعاتی بزرگ می باشد که از طریق آن کلیه داده های حال و گذشته یک سازمان جهت انجام عملیات گزارش گیری و آنالیز در دسترس مدیران قرار می گیرد.
- انبار داده نقش مهمی در تصمیم گیری مدیران سازمان برای تعیین یک راهبرد موفق دارد.
- بعضی از داده ها قبل از ورود به انبار به یک فضای عملیاتی کوچکتر **operational data store** برای پردازش بیشتر وارد می شود.
- انبار داده سه لایه دارد که به ترتیب **Integration**، **Staging** و **Presentation** نام دارد.
- داده های خام از منابع اطلاعاتی مختلفی جمع آوری شده و در لایه **staging** وارد می شوند. منبع داده خام می تواند یک سیستم **ERP**، پایگاه داده یک برنامه کاربردی و یا یک فایل **Excel** باشد.
- ایجاد یکنواختی و جامعیت بین داده های وارد شده به انبار، در دومین لایه یعنی **integration** انجام می شود. به عنوان مثال حذف رکوردهای تکراری و یا نرمال سازی داده ها.
- در لایه **Presentation** داده ها در دسترس کاربران قرار می گیرد. نرم افزارهای تهیه گزارش مانند **Cognos**، **Business Object**، **SAP** با دسترسی به این لایه می توانند اطلاعات مورد نیاز مدیران و تحلیلگران را استخراج و در قالب گزارش یا **Dashboard** عرضه نمایند.
- از آنجائیکه انجام عملیات آماری و گزارشات پیچیده (**OLAP**) دارای بارکاری بسیار سنگینی برای سرورهای پایگاه داده می باشند، وجود انبار داده سبب می گردد که اینگونه عملیات تاثیری بر فعالیت برنامه های کاربردی سازمان (**OLTP**) نداشته باشد.

پشتیبانی بلادرنگ از انبار داده‌های فعال

REAL-TIME SUPPORT FROM ACTIVE DATA WAREHOUSE (ADW)

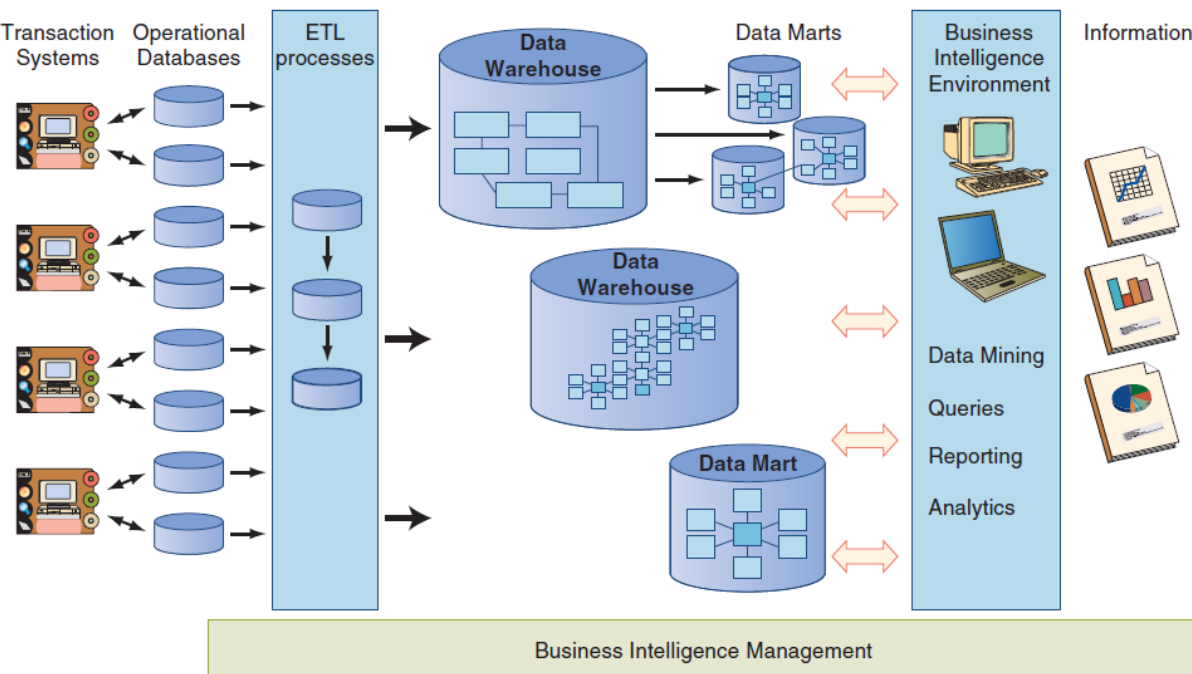
15

- در گذشته نیاز به تحلیل‌های بلادرنگ کمتر بود و نیازی به تعامل مستقیم مشتری یا تجمیع بلافاصله پایگاه داده‌ها نبود. فرآیند ETL (Extract, Transform and Load) یکبار در هفته یا حتی یکبار در ماه انجام می‌گرفت.
- اما امروزه بسیاری از کسب و کارها از اطلاعات در لحظه استفاده می‌کنند تا تعاملی برخط با مشتری داشته باشند.
- شرکتهایی که با ADW کار می‌کنند می‌توانند سرویس‌های بهتری به مشتریان خود بدهند و در مقابل سود بیشتری کسب نمایند.

مثال: مزیت ADW

□ فرض کنید یکی از مشتریان یک شرکت مخابراتی که خدمات مختلفی مانند تلفن همراه، انواع اینترنت، پخش سلولی و غیره ارائه می‌دهد، به علت مشکلی که برایش پیش آمده با نماینده خدمات پشتیبانی شرکت (CSR) تماس می‌گیرد. چون نماینده به ADW دسترسی دارد و کل سابقه مکالمات، سرویس‌ها و خدماتی که مشتری استفاده نموده و امتیاز ارزشی که مشتری برای شرکت دارد را می‌بیند، سعی می‌کند به شکل مناسبی پاسخ آن مشتری را بدهد.

انبار داده (Data Warehouse)



□ علت نیاز به انبار داده:

□ بهبود گزارش‌گیری‌ها

□ درک بهتر کسب و کار

□ شرکت‌های موفق با استفاده خوب از داده‌ها، به تغییرات بازار و فرصت‌ها پاسخ بهتری می‌دهند.

فرآیند ساخت و استفاده از انبار داده:

داده‌های سازمانی روی سیستم‌های عملیاتی ذخیره می‌شوند. لزوماً همه داده‌ها به انبار داده منتقل نمی‌شوند. داده‌هایی که منتقل شده‌اند، به شکلی در انبار داده سازمان‌دهی می‌شوند که دسترسی آنها برای کاربران نهایی ساده باشد.

مزایای انبار داده

18

- برخی حوزه‌های سازمانی که از انبارداری داده‌ها بهره می‌برند:
- بازاریابی و فروش
 - استفاده از انبار داده برای نگهداری اطلاعات مربوط به وضعیت کالاها، میزان تاثیرگذاری برنامه‌های بازاریابی، سود خط تولید و گرفتن تصمیمات هوشمندانه برای حداکثر نمودن سودآوری مشتریان
- قیمت‌گذاری و بستن قرارداد
 - استفاده از داده‌ها برای تعیین دقیق قیمت قراردادها
- پیش‌بینی
 - پیش‌بینی نیازهای کاربران
- کارآیی فروش
 - تعیین میزان سود فروش در مناطق مختلف
- مدیریت مالی
 - استفاده از نتایج روزانه، هفتگی یا ماهانه برای بهبود مدیریت مالی

کاربردهای انبار داده

جدول ۳-۳ کاربردهای استراتژیک انبارداری داده‌ها		
کاربرد استراتژیک	حوزه‌های استفاده کارکردی	صنعت
تخصیص خدمه، استقرار هواپیما، ترکیبی از کرایه‌ها، تحلیل سودآوری مسیرها، تبلیغات برنامه‌های پروازی مکرر	عملیات و بازاریابی	شرکت‌های هواپیمایی
تجارت، تجدید موجودی	توزیع و بازاریابی	پوشاک
خدمات مشتری، تحلیل روند، تبلیغات درباره خدمات و محصولات، کاهش هزینه‌های سیستم‌های اطلاعاتی	توسعه محصول، عملیات و بازاریابی	بانکداری
خدمات مشتری، خدمات اطلاعاتی جدید در قبال دریافت وجه، تشخیص تقلب	توسعه محصول و بازاریابی	کارت اعتباری
انتقال فناوری، تولید برنامه‌های کاربردی نظامی	توسعه محصول	قراردادهای دفاعی
انبارهای داده با قابلیت‌های شخصی‌سازی، ترجیحات بازاریابی و خرید که منجر به Up-Selling و Cross-Selling می‌شوند	توزیع و بازاریابی	کسب و کار الکترونیکی
گزارش‌دهی نواحی ارتکاب جرم، امنیت ملی	عملیات	دولت
کاهش هزینه‌های عملیاتی	عملیات	بهداشت و درمان
مدیریت ریسک، تحلیل حرکت‌های بازار، تحلیل تمایلات مشتریان و مدیریت سبد سرمایه‌گذاری	توسعه محصول، عملیات و بازاریابی	سرمایه‌گذاری و بیمه
تحلیل روند، تحلیل الگوی خرید، سیاست قیمت‌گذاری، کنترل موجودی، تبلیغات فروش، تصمیم‌گیری در مورد کانال‌های توزیع بهینه	توزیع و بازاریابی	زنجیره فروش (خرده‌فروشی)
تبلیغ محصولات و خدمات جدید، کاهش بودجه سیستم اطلاعاتی، تحلیل سودآوری	توسعه محصول، عملیات و بازاریابی	مخابرات

- Data mart شکل کوچک تر و اختصاصی تر انبار داده است.
- پیاده سازی انبار داده هزینه بالایی دارد که برای بسیاری از شرکتها مناسب نمی باشد.
- بنابراین به عنوان یک راه حل جایگزین، نسخه کوچکتر و با هزینه کمتری از انبار داده بنام مارت داده یا بازار داده ایجاد می شود که از نظر هزینه و زمان پیاده سازی به صرفه تر می باشد.

داده کاوی (Data Mining)

- داده کاوی یا کشف دانش، به فرآیند تحلیل داده‌ها از دیدگاه‌های مختلف و خلاصه‌سازی آنها به اطلاعاتی گفته می‌شود که بتوان از آن برای افزایش درآمد و کاهش هزینه‌ها استفاده نمود.
- ابزارهای داده کاوی امکان تحلیل، دسته‌بندی، بررسی همبستگی و الگوهای بین فیلدهای موجود در انبار داده و در نهایت استخراج دانش را فراهم می‌آورند.

چند نرم افزار داده کاوی قدرتمند

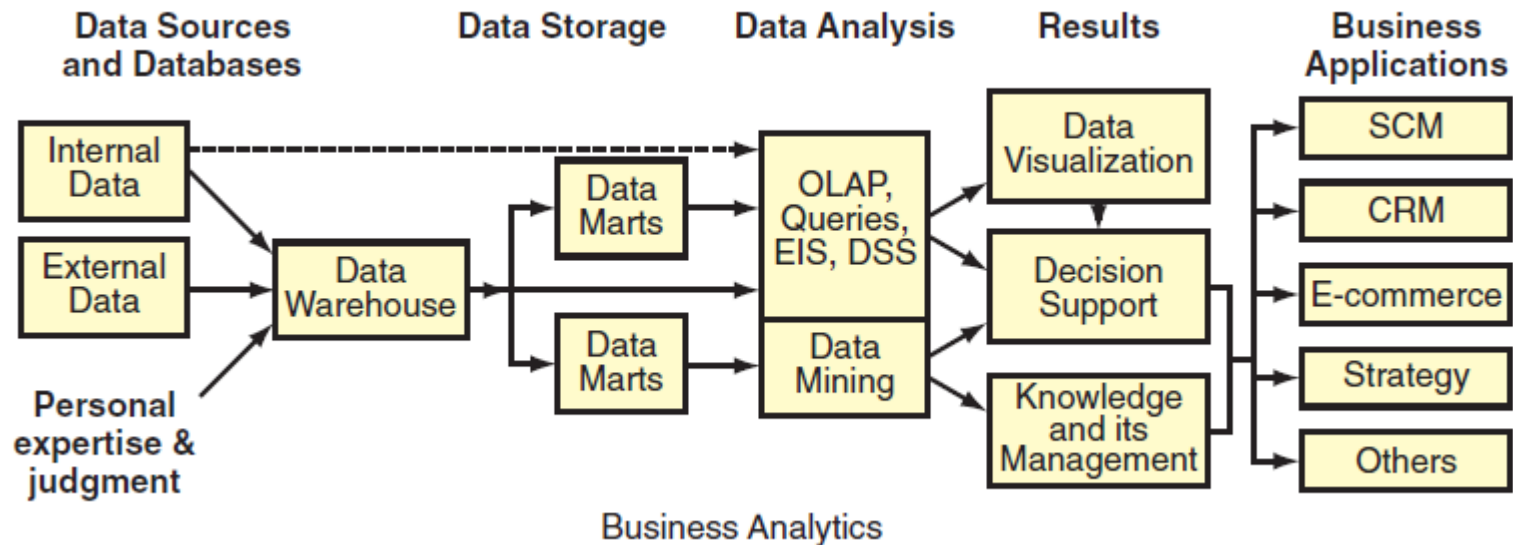


بخش دوم مدیریت اسناد، متون و داده‌ها

چرخه حیات داده (Data Life Cycle)

24

- مدلی است که گردش داده در سازمان را نشان می‌دهد.
- با ذخیره‌سازی داده در پایگاه داده آغاز می‌شود. داده‌ها برای تحلیل، در انبار داده بارگذاری می‌شوند. سپس برای گزارش‌گیری و استفاده در کاربردهای تجاری به کار گرفته می‌شوند.
- مدیریت زنجیره تامین (Supply chain management - SCM)، مدیریت ارتباط مشتری (customer relationship management (CRM) و تجارت الکترونیک کاربردهای سازمانی هستند که برای عملکرد مناسب به داده‌های بروز با دسترسی ساده نیاز دارند.



سه اصل برآمده از چرخه حیات داده

25

□ اصل کاهش ارزش داده‌ها:

■ اصلی ساده اما قدرتمند که می‌گوید با گذشت زمان ارزش داده‌ها کم می‌شود. در واقع با داده‌های موجود اغلب به مدت یک‌ماه می‌توان در اوج بود!

□ اصل ۹۰/۹۰ استفاده از داده‌ها:

■ ۹۰٪ داده‌ها بعد از ۹۰ روز به ندرت مورد استفاده قرار می‌گیرند. نشان دهنده مزیت توانایی انجام اقدام بر اساس داده‌های بلادرنگ

□ اصل داده‌ها در محتوا:

■ داده‌ها باید یکپارچه، پردازش و تحلیل شوند و به صورت اطلاعات قابل تعقیب درآیند. کاربران نهایی، باید داده‌ها را با شکل و ساختاری هدفمند ببینند. بدین منظور می‌توان از ابزارهای داده‌نمایی (visualization) و تصمیم‌یار استفاده نمود.

اهمیت اسناد و متون

26

□ متون دارای قدرتی پنهان در رسیدن به بینش بازاریابی، توسعه محصول جدید، خدمات مشتری، روابط عمومی و رقابت است.

□ بیش از ۷۵٪ داده‌های سازمان‌ها، بدون ساختار هستند

■ شامل اسناد واژه‌پردازی، محتوای اسناد وب، توثیقات و سایر رسانه‌های اجتماعی، ایمیل و پیام‌های متنی، صوت، ویدئو، تصاویر، نمودارها، فکس و یادداشت

■ بنابراین نیاز به روشی برای تحلیل متون داریم:

■ متن کاوی و تحلیل متن



متن کاوی (Text Mining)

27

- روشی برای تحلیل متون که مستلزم تفسیر کلمات و مفاهیم موجود در متن است.
- متن، با هدف آماده‌سازی نکات قابل تعقیب برای مدیران، تحلیل می‌شود.
- تحلیل متون به دو چالش کسب و کار می‌پردازد:
 - ۱- سازمان‌دهی اطلاعات و قابلیت کشف محتوا در اسناد
 - ۲- کشف روندها و الگوهای موردنیاز از اطلاعات متنی برای آینده‌نگری

فرآیند متن کاوی

28

□ بررسی (Exploration)

□ بررسی اسناد برای شمارش کلمات؛ طبقه‌بندی دستی اسناد با مطالعه نمونه‌ای از آنها

□ پیش‌پردازش

□ قبل از تحلیل و دسته‌بندی خودکار محتوا، لازم است پیش‌پردازش‌هایی مانند آماده‌سازی و استانداردسازی، اصلاح غلط املایی، تبدیل کلمات اختصاری، تگ گذاری و غیره انجام شود.

□ دسته‌بندی و مدل‌سازی

□ استفاده از مدل‌های آماری و قوانین کسب و کار برای طبقه‌بندی پیام‌ها و اسناد. آزمودن مدل یادگرفته شده برای اسناد جدید و سپس استفاده انبوه از مدل تولید شده

سیستم مدیریت سند (Document Management System-DMS)

29

□ مدیریت سند

□ کنترل خودکار اسناد تصویری و الکترونیکی، تصاویر صفحات، صفحات گسترده، پیام‌های صوتی و ایمیل، اسناد واژه‌پردازی و دیگر اسناد از ایجاد تا آرشیو و از بین بردن آنها در طی چرخه حیات سازمان، مدیریت اسناد گفته می‌شود.

□ DMS متشکل از سخت افزار و نرم افزارهایی است که اسناد الکترونیکی را مدیریت و آرشیو نموده و اسناد کاغذی را به الکترونیکی تبدیل و شاخص گذاری می‌کند و با توجه به سیاست‌های سازمان آنها را ذخیره می‌کند. مثلاً اسناد مالی ممکن است حداقل هفت سال و ایمیل‌های تبلیغاتی تا یک سال نگهداری شوند.

□ در DMS امکان جستجو و پرس‌وجو وجود دارد.

□ ممکن است DMS بخشی از سیستم یکپارچه مدیریت محتوای سازمانی (ECM) باشد.

سیستم مدیریت سند (Document Management System-DMS)

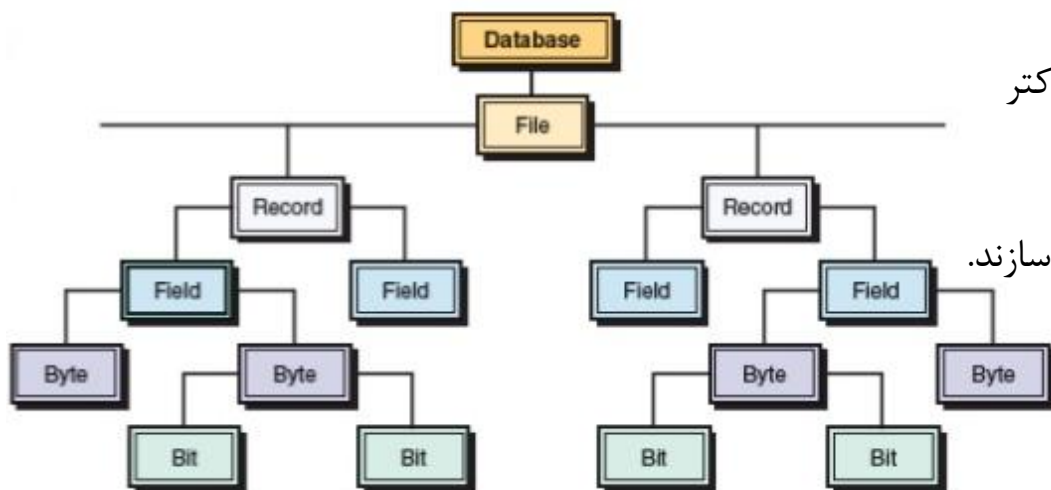
30

- مزایای DMS در کارآیی و بهره‌وری کسب و کار
 - ایجاد قابلیت دستیابی و استفاده از محتویات موجود در اسناد
 - کاهش هزینه نیروی کار با خودکارسازی فرآیندهای کسب و کار
 - کاهش زمان و میزان تلاش لازم برای شناسایی اطلاعات موردنیاز کسب و کار برای پشتیبانی تصمیم‌گیری
 - بهبود امنیت محتوا و در نتیجه کاهش ریسک سرقت مالکیت معنوی
 - کم کردن هزینه‌های چاپ، ذخیره‌سازی و جستجوی محتوا
- ابزارهای اصلی مدیریت سند:
 - نرم افزار گردش کار، ابزارهای تألیف، اسکنرها و پایگاه‌های داده

سیستم‌های مدیریت فایل

سلسله مراتب سازمان دهی داده ها در سیستم کامپیوتری

32



ID No.	Name	D.o.B.	Phone	Class	Tutor	Room
356	Jess	3 Mar 1995	7564356	5B	Mr Noggin	56
412	Hamad	12 Nov 1994	7465846	5B	Mr Noggin	56
459	Sita	9 Jan 1994	8565634	6Y	Ms Take	18
502	Hamad	3 Mar 1995	6554546	5B	Mr Noggin	56

One Record

بیت: کوچک ترین واحد داده که کامپیوتر می تواند آن را پردازش نماید

مثال: ۰ یا ۱

بایت: یک گروه هشت تایی از بیت ها که نمایانگر یک کاراکتر است

مثال: ۰۰۱۰۰۰۱۱ بیانگر عدد ۳۵ در مبنای ده می باشد.

فیلد: مجموعه ای از کاراکترها که یک کلمه یا عدد را می سازند.

به آن صفت هم گفته می شود

مانند: فیلد با عنوان Cust_Name برای نام مشتری

رکورد: مجموعه ای از فیلدهای مرتبط باهم

به آن موجودیت نیز گفته می شود

مثال: ادغام فیلدهای نام فروشنده، آدرس، تلفن، شماره حساب

فایل: مجموعه ای از رکوردهای مرتبط

مثال: رکوردهای همه مشتریانی که در یک موسسه وام مسکن گرفته اند

پایگاه داده: گروهی منطقی از فایل های مرتبط

مثال: همه فایل های وام های مختلف مسکن، خودرو و غیره پایگاه داده وام غیرتجاری را تشکیل دهند.

کلید در پایگاه داده

33

- کلید اصلی: صفت (فیلدی) که برای شناسایی هر رکورد استفاده می‌شود. این فیلد یک شناسه منحصر بفرد برای آن رکورد است.
- کلید ثانویه: فیلدهای غیر منحصر بفردی که مقداری اطلاعات شناسایی دارند.
- کلید خارجی: کلیدهایی که به منظور برقراری ارتباط بین دو یا چند جدول استفاده می‌شوند.

